

24. Los diferentes métodos para evaluar el impacto de una intervención

Andrés Martínez Fernández¹

Evaluar el impacto de una intervención (también de una intervención de desarrollo y más concretamente el impacto de las TIC en la mejora de la salud) significa, no sólo verificar las diferencias encontradas en el colectivo meta, antes y después de la intervención, sino también, asegurarse de que esas diferencias se deben a dicha intervención, lo que no suele ser fácil.

Una dificultad añadida para medir el impacto de un proyecto en una zona rural de un país en desarrollo es que no suele ser fácil encontrar información objetiva, recopilada por otros actores (información secundaria), que mida precisamente el efecto de nuestro proyecto. Además, algunas veces se mide ese efecto de una manera global, y es imposible desagregar esa información sólo para el colectivo meta deseado. Esto tiene una consecuencia importante: nos obliga a recabar nosotros dicha información (información primaria) con el esfuerzo y el coste que esto trae asociado.

24.1. Procedimientos para obtener y procesar información secundaria

Como ya hemos mencionado, la información secundaria es aquella que ha sido obtenida por terceras personas o instituciones. Cualquier evaluación que se base en información dada por otros, siempre que sea de confianza, tiene una validez mayor que la evaluación que se basa en información primaria (la obtenida por nosotros). La posibilidad de realizar una evaluación con información secundaria debe valorarse desde el momento mismo del diseño del proyecto. Unos IOV (indicadores objetivamente verificables) del OG (objetivo general) y del OE (objetivo específico) que puedan ser soportados por FV (fuentes de verificación) externas podrían casi automatizar el proceso de evaluación de impacto. El problema en muchos casos, como ya hemos comentado, es que esas FV externas no existen.

¹Universidad Rey Juan Carlos, España

Queda claro entonces, que el primer paso para saber si existen o no dichas FV externas ha de ser una revisión de las diferentes instituciones, tanto públicas como privadas, que están trabajando en la zona. Una vez identificadas habrá que conocer si están o no obteniendo información sobre alguno de los aspectos relevantes de impacto de nuestro proyecto. Si es así, estudiaremos sus fuentes y la granularidad con la que obtienen la información (para ver si es compatible con la definición de nuestros beneficiarios). Sería muy importante comprobar si cuentan con series temporales históricas (para poder realizar un estudio antes y después) y si también realizan mediciones en otras zonas que pudieran servir como grupo de control (los estudios antes y después con grupo de control presentan una alta validez). Si tenemos esta suerte, y la información de terceros abarca todos los efectos (los impactos) que queremos medir en nuestro proyecto, entonces nos habremos ahorrado mucho trabajo y coste. Aún así, hay que procesar dicha información para saber si las diferencias encontradas antes y después de la intervención, y entre el grupo de aplicación y el grupo de control, son significativas.

Estas herramientas estadísticas son las mismas que habrá que utilizar si la información a utilizar es primaria, por lo que se explicarán en el siguiente apartado.

24.2. Procedimientos para obtener y procesar información primaria

Si no contamos con información externa para verificar el impacto de nuestra intervención, no queda otro remedio que diseñar un procedimiento de medición de determinados indicadores de impacto, para comprobar si han variado significativamente antes y después del proyecto.

La experiencia nos hace saber que no se puede medir nada que no se conozca previamente. Con esto se quiere decir que no es razonable construir un instrumento de medida (una herramienta cuantitativa) sin que antes conozcamos de forma cualitativa, si ha habido o no alguna diferencia importante entre el antes y el después de la intervención.

24.2.1. Las herramientas cualitativas

La investigación cualitativa puede ser vista como una investigación en sí misma (ha demostrado que puede ofrecer la información necesaria para entender la explicación de un fenómeno social en un tiempo y a un coste muy reducidos), o puede verse como la base de una posterior investigación cuantitativa (como es el caso que nos ocupa) con el objetivo de poder afinar mucho más las herramientas de medición, con la información que ha aportado sobre la variable en cuestión.

Las principales herramientas cualitativas son cuatro:

Las reuniones de grupo. Son reuniones de tamaño mediano, en las que están representadas las diferentes posturas o actores de un tema en cuestión, que conversan

o discuten de un problema concreto, conducidos por un moderador con un plan previo. El objetivo en este caso es la extracción de información acerca de las diferentes teorías o pensamientos sobre los resultados del proyecto y el impacto que está produciendo. No se pretenden alcanzar consensos o entender en profundidad cada una de las cuestiones aparecidas, sino identificar los diferentes aspectos clave, actores principales, fuentes de información, etc. Se suelen grabar en vídeo (con consentimiento de los presentes) para un posterior análisis en profundidad.

Las entrevistas en profundidad. Tras la identificación en la reunión de grupo, de los aspectos clave (en nuestro caso los impactos más importantes) o de las personas que tienen más información sobre el proyecto, se pueden preparar entrevistas en profundidad con ellas. Las entrevistas en profundidad son reuniones para intercambiar información detallada entre el entrevistador (investigador) y el entrevistado (experto, afectado por el proyecto en cuestión, etc.). Estas entrevistas pueden ser estructuradas (con preguntas muy concretas), semiestructuradas (con los temas a tratar bien identificados) o abiertas (conversación extensa sobre el tema a tratar). Deberían ser grabadas o al menos tomar buenas notas de la misma.

La observación participante. La observación participante no es únicamente la contemplación del entorno (en nuestro caso, los establecimientos de salud o la situación de los beneficiarios), sino adentrarse en él hasta entenderlo. Requiere un papel activo, con reflexión permanente acerca de lo que observamos. Su objetivo también puede ser múltiple; en nuestro caso intentaremos comprender procesos y hechos relevantes en los que el proyecto ha tenido efectos, entender la interacción y la vinculación de los diferentes actores (personal de atención y pacientes), identificar impactos, etc. La observación participante suele ser lenta y no siempre puede llevarse a cabo (permisos, conflictos, costes, etc.). Esta técnica exige el uso asiduo de un cuaderno de campo.

La revisión de textos y documentación. Nos referimos a cualquier tipo de documento que pueda aportar información relevante (folletos, notas de reuniones, acuerdos, memorias de actividad, registros contables, etc.). Tiene la ventaja de obtenerse sin reacción a la presencia del observador, pero por el contrario, podría hallarse manipulado, filtrado, alterado, con un sesgo hacia los intereses de la persona o colectivo encargados de su preparación. Suele ofrecer gran cantidad de información a un coste relativamente bajo.

La validez de los resultados obtenidos con herramientas cualitativas es cuestionada en muchas ocasiones. Los prejuicios del investigador, sus intereses, percepciones, observaciones, conocimientos y su actitud crítica desempeñan un papel importante en el estudio. Para la validación se recomiendan tres herramientas prácticas:

La triangulación. El uso de múltiples fuentes de datos suele aumentar la solidez y la validez de los resultados. El empleo de más de un método de obtención de datos permite a menudo encontrar relaciones para sugerir o contrastar hipótesis. Asimismo, la revisión por más de un investigador ayuda a eliminar subjetivismos y a detectar imprecisiones e incorrecciones.

La visión global. El resultado de una investigación cualitativa debería permitir comprender globalmente el sistema o la situación problemática. Si esto no se produce, puede ser debido a que no se hayan recabado datos suficientes o a que la perspectiva adoptada o la metodología utilizada sean muy parciales o incorrectas.

La realimentación. Antes de pasar a la etapa de validación cuantitativa, conviene contrastar con los diferentes actores el resultado del proceso de investigación cualitativa. Si el personal de atención de salud o los beneficiarios del proyecto (en el caso que nos ocupa) no entienden o no comparten los resultados de impacto obtenidos, tal vez convendría profundizar más o volver a replantear las hipótesis de partida de la investigación.

24.2.2. Las herramientas cuantitativas

Una vez conocidos los impactos del proyecto de manera cualitativa, queda contrastar la magnitud de los mismos. Para ello, se utilizan las mismas herramientas metodológicas que en la preparación de un experimento científico. Es importante resaltar en este momento, que el proceso de evaluación debe estar planteado antes de iniciado el proyecto, ya que *a posteriori* es difícil, por no decir en muchos casos imposible, acceder a información necesaria de lo que se viene a llamar “línea de base” (situación de partida, previa a la intervención).

Si definimos un experimento (prospectivo) como una situación controlada (ya hablaremos de cómo controlar las variables extrañas) en la que el investigador manipula intencionalmente una variable independiente (en nuestro caso, la introducción de las TIC) para medir su efecto sobre una variable dependiente (por ejemplo, la mortalidad materna o el número de transferencias urgentes entre establecimientos de salud), vemos que la evaluación de impacto de un proyecto no deja de ser un experimento. Un experimento se caracteriza por la existencia de dos grupos, uno de intervención y otro de control, que han sido definidos al azar, sobre los que sí y no, respectivamente, se manipula la variable independiente, pero en ambos se mide la variable dependiente.

Es importante distinguir un experimento de un estudio cuasiexperimental (no hay grupo de control, o la asignación de elementos muestrales al grupo de tratamiento o de control no ha sido aleatoria). Eso nos puede ocurrir en muchos de nuestros proyectos piloto (en muchos casos es el Ministerio de Salud quien prioriza los establecimientos a los que instalar las TIC). En este caso, la evidencia aportada es un poco menor que en un experimento, pero aún así valiosísima si se hace bien.

Un aspecto clave a tener en cuenta al diseñar un experimento (una evaluación de impacto en nuestro caso) es la validez de los resultados que vamos a obtener. Se suele distinguir entre validez interna y validez externa.

La validez interna trata de responder a la pregunta de si el tratamiento experimental de la variable independiente (en nuestro caso, la introducción de un determinado sistema de telemedicina) es realmente la causante de la modificación observada en la variable dependiente (por ejemplo, la reducción en mortalidad materna). La validez interna requiere controlar adecuadamente las “variables extrañas”, que son factores externos

al experimento que pudieran distorsionar o influir sobre los resultados (por ejemplo, un cambio de política en la medición de la mortalidad materna directa o indirecta, o la aplicación en el mismo momento de un presupuesto adicional para luchar contra la mortalidad materna en todo el país, etc). Aunque existen muchos procedimientos para mejorar la validez interna, el más utilizado es el de la introducción de un grupo de control. Si existen variables espurias, afectarán por igual al grupo de intervención que al de control; por lo tanto, midiendo la diferencia entre ambos grupos podremos obtener el impacto que se ha debido únicamente a la intervención.

La validez externa es otra cosa. Trata de responder a la pregunta de si son generalizables los resultados de la investigación a otros escenarios menos controlados. Por ejemplo, si he realizado el proyecto en establecimientos de atención de salud de una ONG en donde hay un control muy exhaustivo del personal, ¿serán aplicables los resultados a los establecimientos del Ministerio de Salud?

Este capítulo no puede convertirse en un tratado de estadística, pero sí que puede explicar los conceptos fundamentales y presentar algunos ejemplos muy sencillos, que nos podrían ayudar a entender los procedimientos más usados en la evaluación de impacto de proyectos de telemedicina.

24.2.2.1. La medición directa o la encuesta

Ya hemos explicado que para medir el impacto de una determinada intervención, debemos definir las variables que creemos pueden cambiar debido al proyecto (la investigación cualitativa nos ofrecerá cuáles pueden ser) y éstas deben ser “medidas” antes y después del proyecto. Algunas de esas variables pueden ser medidas automáticamente (por ejemplo, número de llamadas telefónicas al centro de salud de referencia en el último mes, que podría quedar registrado en la centralita) pero otras sólo pueden obtenerse encuestando a una determinada población (a una muestra del personal de atención o a una muestra de pacientes, por ejemplo). Diseñar una buena encuesta no es tarea fácil.

Una encuesta es un conjunto de preguntas normalizadas (agrupadas en un cuestionario) dirigidas a una muestra representativa de la población (si es a toda la población se llama censo) o instituciones, con el fin de conocer estados de opinión o hechos específicos. Tiene la ventaja de permitir el acceso a muchos individuos a un coste “razonable”, permitir cuantificar aspectos humanos o psicológicos (como opiniones o pensamientos), aunque a veces puede presentar problemas de validez externa puesto que en determinados temas, mucha gente no contesta la verdad.

En una encuesta utilizamos un instrumento de medida (el cuestionario) que en muchos casos no existía, y no estaba probado (es como construir tu propio voltímetro para realizar una medida de tensión eléctrica en un experimento). Antes de empezar a medir, hemos de validar el instrumento de medida. Tenemos que obtener la “validez” y “consistencia” del cuestionario antes de encuestar, a través de una prueba con una submuestra pequeña.

Es muy importante formar adecuadamente a los encuestadores, y mostrar sensibilidad

hacia el encuestado. El cuestionario debe contener instrucciones claras. Debe comen- zarse con preguntas sencillas y no deben hacerse preguntas innecesarias o injustificadas.

Las variables a medir en un cuestionario pueden obtenerse con una única pregunta (uso del sistema de teleecografía con la pregunta “cuántas teleecografías se han reali- zado en el último mes en su establecimiento”) o con varias (la calidad del servicio de videoconferencia podría obtenerse preguntando por el número de llamadas fallidas en la última semana, el nivel de ruido en la comunicación, etc.).

Las preguntas en un cuestionario suelen clasificarse como “cerradas” (con opciones de respuestas predefinidas) o “abiertas” (informativas, pero más difíciles de procesar). Cuanto más profunda ha sido la investigación cualitativa, más preguntas cerradas podré formular. Las preguntas cerradas pueden clasificarse a su vez en preguntas de alter- nativa simple (sólo se puede elegir una de las respuestas) o de múltiples alternativas (pueden seleccionarse varias respuestas, lo cuál también complica algo su postproce- sado). La codificación previa de las preguntas es importante, y dependerá del tipo de variable:

- Cualitativa. Categorías sin orden (profesión, tipo de establecimiento de salud, etc.).
- Ordinal. Categorías ordenadas (satisfacción con el sistema de telemedicina baja, media, alta, muy alta).
- Cuantitativa. Que a su vez podemos clasificar como continuas (nivel de glucemia de las gestantes), o discretas (cuántos días ha estado estropeado el sistema).

Para saber cuántas medidas hay que hacer, o a cuántas personas hay que entrevistar, introducimos el siguiente apartado. Además, adelante se ofrece una base para realizar un correcto procesamiento de los datos obtenidos.

24.2.2.2. El cálculo del tamaño muestral

El objetivo perseguido en el cálculo del tamaño muestral es encontrar el número mí- nimo de unidades muestrales (n) que necesito para conformar una muestra con cierta representatividad de la población (N).

El tamaño muestral (n) depende de muchos factores (principalmente del tamaño de la población, de la variabilidad de la característica a estudiar en la población, de la precisión con la que quiero dar mi estimación, y del riesgo que estoy dispuesto a asumir por equivocarme).

La fórmula más utilizada para el cálculo del tamaño muestral es:

$$n_0 = \frac{k^2 \times \text{varianza}}{(\text{error de precisión})^2}$$

Donde k : nivel de confianza.

Si se quiere afinar más en función del tamaño de la población (N), entonces el tamaño muestral n es:

$$n = \frac{n_0}{1 + (n_0/N)}$$

k es una constante que depende del nivel de confianza o probabilidad de que los resultados de nuestra investigación sean ciertos. Un nivel de confianza del 95 % ($k=1,96$) significa un riesgo de equivocación de un 5 %. Valores típicos de k son: 1,15 para el 75 % de confianza (25 % de riesgo de equivocación); 1,65 para el 90 % de confianza; 1,96 para el 95 % de confianza; y 2,58 para el 99 % de confianza (un riesgo de equivocación de sólo el 1 %).

Si estamos calculando el tamaño muestral para un porcentaje, la varianza se calcula como $(p \times q)$, siendo p el porcentaje que posee la característica a estudiar y q el que no la tiene. Para diseñar, si no se tiene información previa, habría que poner máxima variabilidad (caso peor), que para un porcentaje es 50 % y 50 %, luego $p \times q = 0,5 \times 0,5 = 0,25$.

Existen muchas páginas web que ayudan a este cálculo, pero para entenderlo, nada mejor que un par de ejemplos sencillos.

Ejemplo 1: El servicio de teleecografía ha atendido a 12.000 pacientes ($N = 12.000$). Quiero saber si han quedado o no satisfechos con la atención. ¿A cuántos tengo que preguntar (n) si quiero tener una seguridad del 95 % ($k = 1,96$) y no quiero tener un error mayor a ± 3 % (error de precisión = 0,03)?

$$n_0 = \frac{1,96^2 \times 0,25}{0,03^2} = 1.067$$

$$n = \frac{1.067}{1 + (1.067/12.000)} = 980$$

Como vemos, no son pocas las encuestas que tengo que hacer. Tendré que valorar si realmente mi riesgo de equivocación necesita ser tan bajo (el nivel de satisfacción no es tan grave), si puedo permitirme un error mayor al 3 %, o incluso si no sería interesante hacer un piloto con 20 usuarios para saber si la satisfacción está cerca o no del 50 %.

Si acepto una seguridad del 90 % ($k = 1,65$), un error de ± 5 % (error de precisión = 0,05) y en la prueba piloto me salía que el 80 % estaban satisfechos ($p \times q = 0,8 \times 0,2 = 0,16$).

$$n_0 = \frac{1,65^2 \times 0,16}{0,05^2} = 174$$

$$n = \frac{174}{1 + (174/12.000)} = 172$$

Vemos que las diferencias son significativas y eso puede afectar mucho al coste de la investigación.

Este cálculo se ha realizado para una proporción. Un ejemplo para el cálculo de medias podría ser el siguiente:

Ejemplo 2: Se desea conocer el coste promedio de las transferencias urgentes de pacientes en todos los establecimientos de salud de Nicaragua (incluyendo zonas bien y mal comunicadas). Se hizo un trabajo inicial de acompañamiento en 25 transferencias urgentes en establecimientos elegidos al azar, comprobando que el coste promedio era de 880 Córdobas, con una desviación típica de 330 Córdobas. Se quiere calcular el tamaño muestral necesario para estimar el coste promedio en todo Nicaragua con un error de precisión de ± 50 Córdobas y un nivel de confianza del 95 % (riesgo del 5 %).

$$n_0 = \frac{1,96^2 \times 330^2}{50^2} = 167$$

Este número se podría reducir un poco si supiéramos el número de transferencias urgentes que se realizan en todo Nicaragua, pero tampoco bajaría tanto.

24.2.2.3. El modelo gaussiano de probabilidad

Los modelos de probabilidad son funciones matemáticas que sirven para emular el comportamiento de determinadas características de poblaciones. De entre todos los modelos de probabilidad que existen, uno sobresale especialmente, el modelo *normal* o *gaussiano*. Son muchísimas las variables que responden a este modelo (calificación en una prueba, tiempo dedicado a realizar una consulta, peso de recién nacidos, etc.) siendo simétricas con respecto a un valor central, alrededor del cual toman valores con mucha probabilidad y en las que apenas aparecen valores extremos.

La distribución normal queda caracterizada por dos parámetros $N(\mu, \sigma)$, donde el primero es la media y el segundo la desviación típica o estándar (valor en el que cambia la concavidad de la curva).

Como podemos ver en la Figura 24.1, una variable gaussiana presenta el 68,3 % de sus valores en el intervalo comprendido entre la media ± 1 desviación típica, el 95,5 % de sus valores en el intervalo comprendido entre la media ± 2 desviaciones típicas, y el 99,7 % en el intervalo comprendido entre la media ± 3 desviaciones típicas.

Aunque no se puede apreciar en la figura, el 95 % de sus valores se encuentran en el intervalo comprendido entre la media $\pm 1,96$ desviaciones típicas, y el 99 % en el intervalo comprendido entre la media $\pm 2,58$ desviaciones típicas (valores muy utilizados en este tipo de experimentos).

Si una variable sigue una distribución gaussiana, se pueden aplicar multitud de técnicas estadísticas muy interesantes, luego será muy importante, antes de nada, comprobar si una muestra realmente obedece a este tipo de distribución.

Para hacerlo se aplicará lo que en estadística se conoce como contraste de ajuste de modelos. No vamos a entrar en mucho detalle, pero consistirá en comparar el

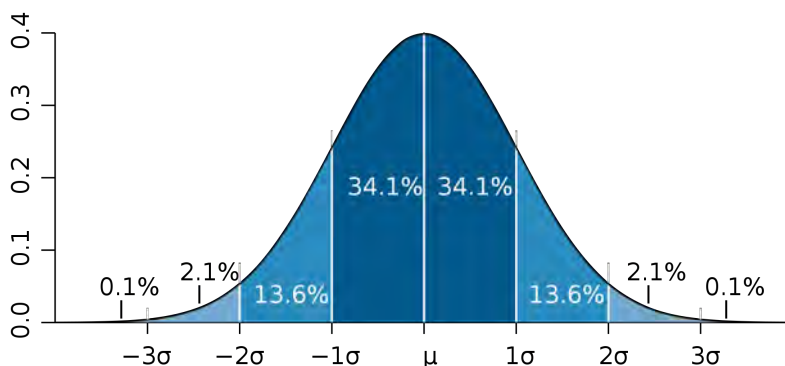


Figura 24.1.: Modelo Gaussiano. Fuente: M. Toews. Licencia CC BY 2.5. Disponible en: http://commons.wikimedia.org/wiki/File:Standard_deviation_diagram.svg.

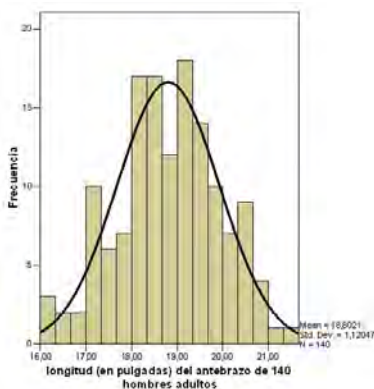


Figura 24.2.: Histograma y campana de Gauss.

histograma de la variable con la función gaussiana correspondiente (Figura 24.2). Si la diferencia entre las dos gráficas no es mucha podremos asumir que esa variable es gaussiana. Si las diferencias son notables, no podremos hacerlo y tendremos que tener cuidado para aplicar algunos procedimientos estadísticos reservados para este tipo de variables.

Existen básicamente dos pruebas para establecer la “normalidad” o “gaussianidad” de una variable: el contraste chi-cuadrado y la prueba de Kolmogorov-Smirnov. Ambos métodos aparecen en cualquiera de los paquetes informáticos de estadística.

24.2.2.4. El establecimiento de intervalos de confianza para la media

Una vez aclarado que nuestra muestra responde a un modelo (por ejemplo gaussiano), debemos encontrar los parámetros de dicho modelo (en el caso gaussiano, la media

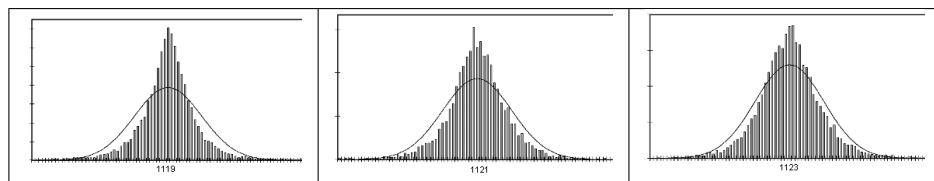


Figura 24.3.: Muestras de población.

poblacional o esperanza, y la desviación típica poblacional). Pero no hemos de olvidar que esos parámetros poblacionales los vamos a obtener de datos de una “muestra concreta”, por lo que no van a ser los correctos. Esta es la razón por la que introducimos el concepto de “estimación por intervalos”. Vamos a dar un intervalo en el que con mucha seguridad o confianza, se encontrará la media poblacional o la desviación típica poblacional. Vamos a explicar con cierto detenimiento la determinación del intervalo de confianza (IC) para la media de una población. La determinación del IC para una proporción o para la desviación típica de una población seguirá procedimientos parecidos, pero no vamos a contarlos en detalle.

En la Figura 24.3, podemos observar que si tomamos diferentes muestras (incluso con el mismo n) de una población normal de media μ y desviación típica σ , representada por $N(\mu, \sigma)$, obtenemos medias muestrales muy parecidas, pero no iguales. ¿Con cuál nos quedamos?

Se puede demostrar que, curiosamente, las diferentes medias muestrales de una población normal (incluso también no normal si $n > 30$) se distribuyen también según una distribución gaussiana, con la misma media y cuya desviación típica disminuye según aumente n . Por lo tanto, cada media muestral que obtengamos seguirá una distribución del tipo:

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Si eso es así, según lo estudiado en el apartado 24.2.2.3, con un 95 % de seguridad, la media muestral (\bar{x}) tiene que cumplir que:

$$\mu - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

Y pasando términos de un lado al otro de la desigualdad, tenemos que:

$$\bar{x} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

Ya hemos aprendido a confeccionar el intervalo de confianza al 95 % para la media de una población.

Lógicamente, si queremos el intervalo de confianza al 99 % de seguridad (si se quiere con otros porcentajes hay que visitar las tablas de gaussianidad):

$$\bar{x} - 2,58 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2,58 \cdot \frac{\sigma}{\sqrt{n}}$$

Nada mejor que un ejemplo para entenderlo. Si se ha tomado una muestra de tamaño 25 de una población normal $N(\mu, 3)$ y la media obtenida en esa muestra es de valor 100, ¿cuál es el intervalo en el que, con una confianza del 95 %, podremos encontrar a la media poblacional μ ?

$$\mu \in \left(100 - 1,96 \cdot \frac{3}{\sqrt{25}}, 100 + 1,96 \cdot \frac{3}{\sqrt{25}} \right) = (98,8, 101,2)$$

De forma genérica, si quisiera el intervalo al $(1 - \alpha) \cdot 100$ % de confianza (a α se le llama el nivel de significación), quedaría construido de la siguiente manera:

$$\mu \in \left(1 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, 1 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Donde $z_{\alpha/2}$: Valor crítico

El valor de $z_{\alpha/2}$ habría que buscarlo en las tablas de la normal.

De todas formas, no podemos “cantar victoria”. Es muy muy raro que si de una población no sabemos la media, sepamos la desviación típica ($iN(\mu, 3)$?). Para solucionarlo, podemos hacer una estimación de la desviación típica poblacional utilizando la desviación típica de la muestra “corregida”, que se calcula como:

$$\hat{S} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}}$$

Resulta ahora que las medias muestrales calculadas de esta manera no siguen una distribución normal, sino una muy parecida, llamada *t de Student* con $n - 1$ grados de libertad. Al igual que la normal, también se encuentra tabulada (es muy fácil de encontrar en Internet).

De nuevo un ejemplo. Si se ha tomado una muestra de tamaño 25 de una población normal con media y desviación típica desconocidas ($N(\mu, \sigma)$) y la media obtenida en esa muestra es de valor 100 y la desviación típica (corregida) es 3, ¿cuál es el intervalo en el que, con una confianza del 95 %, podremos encontrar a la media poblacional μ ?

$$\mu \in \left(100 - t_{(\alpha/2, n-1)} \cdot \frac{3}{\sqrt{25}}, 100 + t_{(\alpha/2, n-1)} \cdot \frac{3}{\sqrt{25}} \right)$$

$$\mu \in \left(100 - t_{(0,05/2, 25-1)} \cdot \frac{3}{5}, 100 + t_{(0,025, 24)} \cdot \frac{3}{5} \right)$$

$$\mu \in \left(100 - 2,064 \cdot \frac{3}{5}, 100 + 2,064 \cdot \frac{3}{5} \right)$$

$$\mu \in (98,8, 101,2)$$

24.2.2.5. El contraste de hipótesis

Llamamos “hipótesis estadística” a una afirmación respecto a una característica de una población. Contrastar una hipótesis es comparar las predicciones que se deducen de ella, con la realidad que observamos; si hay coincidencia (dentro de un margen de error admisible), entonces mantendremos la hipótesis; en caso contrario, la rechazaremos (lo que debería implicar sustituirla por otra hipótesis capaz de explicar los datos observados).

En una investigación, normalmente plantearemos tanto una *hipótesis nula* (H_0) (aquella que refleja la situación actual del conocimiento) como una *hipótesis alternativa* (H_1) (proposición que el investigador aventura con base en sus descubrimientos). La investigación tratará en primer lugar de verificar el *rechazo* de la hipótesis nula, dando paso así a la hipótesis alternativa. La hipótesis nula se presume verdadera hasta que una evidencia estadística (a través de un experimento) muestra lo contrario.

Un ejemplo: un investigador quiere probar si una nueva metodología de aprendizaje basada en el uso de las TIC mejora el aprendizaje de la anatomía humana en las escuelas de enfermería de la India. Para ello selecciona al azar 10 escuelas de enfermería en las que se introducirán las TIC y se aplicará la nueva metodología, y otras 10 (grupo de control) donde se seguirá el método tradicional. Se realiza un examen antes de la introducción de las TIC y otro después.

- Hipótesis nula: No habrá diferencias significativas entre la variación de las notas de ambos grupos.
- Hipótesis alternativa: La introducción de la nueva metodología hace aumentar significativamente los conocimientos de anatomía en los alumnos de las escuelas de enfermería de la India.

Para introducir el concepto de zona de aceptación y rechazo de la hipótesis nula, vamos a plantear un ejemplo: unos investigadores aseguran haber encontrado un nuevo aditivo para la obtención de un pegamento extrafuerte para prótesis dentales, que manteniendo el coste, mejora sensiblemente la resistencia del mismo.

Sabíamos que el pegamento tradicional tenía unas propiedades de resistencia que seguían una distribución $N(5.000, 200)$. Para comprobar que la afirmación es realmente cierta (aumento de resistencia con mantenimiento de costes) se pegan 50 piezas con la nueva fórmula y se verifica que la media muestral de la resistencia es ahora 5.220.

La hipótesis nula en nuestro caso (H_0) es que no ha habido variación de la media, es decir que:

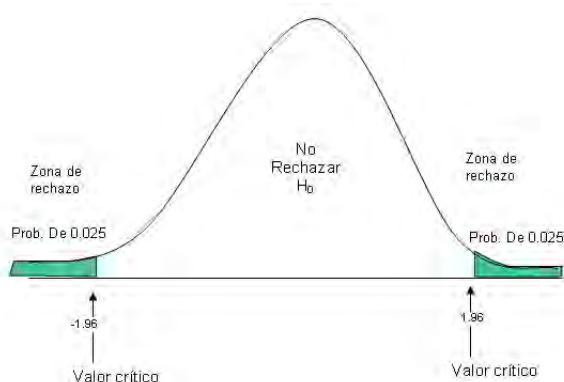


Figura 24.4.: Zonas de aceptación y rechazo de la hipótesis nula.

$$\bar{x} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Si queremos comprobar dicha hipótesis con un 95 % de confianza (significación α del 0,05), tendremos que verificar que (Figura 24.4):

$$\bar{x} \in \left(5.000 - 1,96 \cdot \frac{200}{\sqrt{50}}, 5.000 + 1,96 \cdot \frac{200}{\sqrt{50}}\right) = (4.944,6, 5.055,4)$$

$$\bar{x} = 5.220 \notin (4.944,6, 5.055,4)$$

Diremos que no aceptamos H_0 (que sería que no ha habido variación) con un nivel de significación del 5 %. Lo que estamos diciendo ahora es que con una probabilidad de equivocarme sólo del 5 %, me atrevo a decir que este nuevo aditivo ha modificado las condiciones de resistencia del pegamento (lo cuál sería un éxito para los investigadores).

Pero ¿y no podría decirlo con una significación aún mayor? Eso daría más contundencia. En este caso superamos incluso un α de 0,001 (confianza del 99 % o probabilidad de equivocarme menor del 1 %). Estaría bien que calculáramos el denominado “p valor”, llamado a veces nivel crítico de contraste, que no es más que la mínima probabilidad de equivocarme al decir que rechazo la hipótesis nula. Podría decir que estos investigadores han encontrado algo bueno, con un riesgo de equivocarme de $p \cdot 100$ %.

Este mismo ejemplo nos sirve para introducir otro concepto importante; el concepto de contraste unilateral o bilateral (Figura 24.5). Hay que darse cuenta de que si lo que realmente me interesa demostrar son sólo mejoras, los valores a la izquierda de la media están más a favor de H_0 (no hay mejora en la resistencia) que de H_1 (he encontrado algo que mejora la resistencia). Si eso es así, las zonas de rechazo y de aceptación de H_0 podrían variar y trabajar no con $z_{\alpha/2}$ sino con z_{α} , haciendo un contraste unilateral a la derecha.

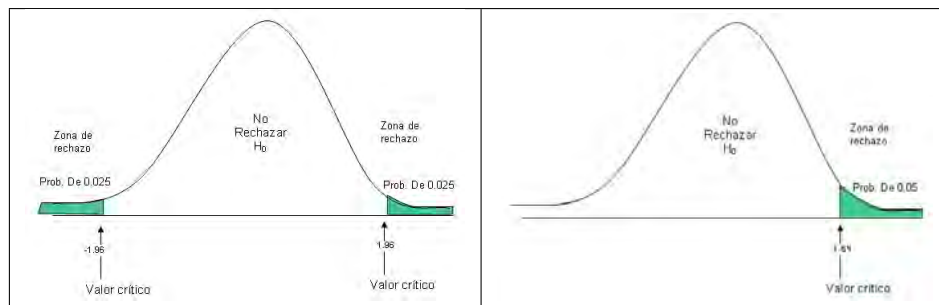


Figura 24.5.: Contraste bilateral vs unilateral.

Por desgracia, este ejemplo anterior no es fácil de encontrar en un estudio de impacto de un proyecto de telemedicina. Hay que darse cuenta de que lo normal no es conocer la distribución de la población y comparar los resultados de mi muestra con ellos; lo normal va a ser elegir una muestra antes de modificar la variable independiente y luego elegir otra muestra después de modificar la variable independiente, y con eso quiero saber si hay o no hay diferencias significativas en las medias. Vamos a proceder a explicarlo con un último ejemplo.

Queremos saber si el uso de un computador ahorra tiempo de trabajo a la hora de rellenar los informes epidemiológicos en los establecimientos de salud de Cuba. Para ello elegimos al azar a 10 trabajadores que no usan computadora y a otros 10 que sí usan computadora. Las horas dedicadas mensualmente por cada uno de ellos han sido:

SinComp	ConComp
51,3	29,6
39,4	47,0
26,3	25,9
39,0	13,0
48,1	33,1
34,2	22,1
69,8	34,1
31,3	19,5
45,2	43,8
46,4	24,9

Para la muestra número 1 (sin computadora) obtenemos una media muestral de valor $\bar{x}_1 = 43,1$ y para la segunda muestra (con computadora) de valor $\bar{x}_2 = 29,3$. Además, la desviación típica de la primera muestra es $S_1 = 12,3$ y de la segunda, $S_2 = 10,6$. Con lo que ya hemos estudiado podríamos hacer el intervalo de confianza para cada una de las muestras y ver si se mezclan, pero estamos perdiendo información que podríamos utilizar por el hecho de tener una muestra de 20 casos en vez de dos de 10 casos.

Lo que sabemos es que si ambas muestras provienen de poblaciones gaussianas con la misma variabilidad (homocedasticidad), variarán únicamente en la media. La H_0 , que es que las medias son iguales, se puede expresar también diciendo que la diferencia de medias seguirá una distribución:

$$N\left(0, \frac{\sigma}{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}\right)$$

Pero claro, al desconocer de nuevo σ , tendremos que utilizar una estimación de la desviación típica poblacional \hat{S}_R :

$$\hat{S}_R = \sqrt{\frac{(n_1 - 1) \cdot \hat{S}_1^2 + (n_2 - 1) \cdot \hat{S}_2^2}{n - 2}}$$

Ya sabemos que si utilizamos \hat{S}_R en vez de σ , la distribución no será una normal sino una *t de Student* de $n - 2$ grados de libertad ($n = n_1 + n_2$). Por lo tanto, la zona de aceptación de H_0 será tal que:

$$\bar{x}_1 - \bar{x}_2 \in \left(0 - t_{(\alpha/2, n-2)} \cdot \frac{\hat{S}_R}{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}, 0 + t_{(\alpha/2, n-2)} \cdot \frac{\hat{S}_R}{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}\right)$$

Para el ejemplo:

$$\bar{x}_1 - \bar{x}_2 \in \left(0 - 2,1 \cdot \frac{11,44}{\sqrt{\frac{100}{20}}}, 0 + 2,1 \cdot \frac{11,44}{\sqrt{\frac{100}{20}}}\right)$$

$$\bar{x}_1 - \bar{x}_2 = 43,1 - 29,3 = 13,8 \notin (-10,74, 10,74)$$

Por lo tanto, podemos afirmar que se rechaza la hipótesis nula con un 95 % de confianza, luego existe una reducción significativa (significación mayor al 0,05) en el tiempo dedicado a rellenar los informes epidemiológicos gracias al uso de computadora.

No podremos decir lo mismo si queremos significación del 0,01 (seguridad del 99 %), ya que en ese caso:

$$\bar{x}_1 - \bar{x}_2 = 13,8 \in \left(0 - 2,88 \cdot \frac{11,44}{\sqrt{\frac{100}{20}}}, 0 + 2,88 \cdot \frac{11,44}{\sqrt{\frac{100}{20}}}\right) = (-14,74, 14,74)$$

Aún así podemos hilar más fino, ya que hemos utilizado el contraste bilateral. ¿Podríamos haber utilizado un contraste unilateral?

$$\bar{x}_1 - \bar{x}_2 \in \left(-\infty, 0 + t_{(\alpha, n-2)} \cdot \frac{\hat{S}_R}{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}} \right)$$

$$\bar{x}_1 - \bar{x}_2 \in \left(-\infty, 0 + 2,55 \cdot \frac{11,44}{\sqrt{\frac{100}{20}}} \right)$$

$$\bar{x}_1 - \bar{x}_2 = 13,8 \notin (-\infty, 13,05)$$

Por lo tanto, sí podemos afirmar que se rechaza la hipótesis nula con un 99 % de confianza, luego existe una reducción significativa (significación mayor al 0,01) en el tiempo dedicado a rellenar los informes epidemiológicos gracias al uso de computadora.

Aunque en este capítulo no hemos podido, ni creemos que deberíamos, explicar muchos conceptos estadísticos importantes, creemos que es más que suficiente para realizar estudios básicos de impacto. Además, nos servirá para entender resultados de investigación dados en revistas científicas. Por último, serán una buena base para empezar a entender libros sobre metodología de investigación, que en muchos casos dan estos conceptos por sabidos.