Decisión estadística mediante la prueba de significación de la hipótesis nula

Vicente Manzano Arrondo – 2009-2014

Una de las expresiones que más impresionan a quienes se acercan a estudiar análisis de datos (sea por voluntad propia o, lo que resulta más frecuente, porque no tienen más remedio) es la que encabeza este documento. Al finalizar su lectura seremos capaces de entender cada uno de sus términos, espero también que sepamos sacarle partido y conocer sus limitaciones. Las pruebas de significación de la hipótesis nula (PSHN) constituyen el método estadístico más utilizado para tomar decisiones sobre si tienen lugar efectos, cambios o relaciones en el ámbito de las ciencias humanas, sociales y de la salud. El objetivo de PSHN es facilitar la toma de decisiones con respecto a la existencia de hipótesis de investigación que permiten ser respondidas con un sí o un no. Algunos ejemplos son ¿Ha cambiado la actitud frente al machismo? ¿Existe relación entre la depresión y la edad? ¿Hay en Europa menos sensibilidad musical que en Asia?

La PSHN ha recibido muchas críticas. Pero no conocerla implica no entender buena parte de las publicaciones que tienen lugar en las ciencias mencionadas sean con carácter específicamente científico, académico o profesional.

De qué hipótesis estamos hablando

Las hipótesis son ideas previas, sentencias que se refieren a algo sobre lo que podríamos ejercer algún tipo de comprobación, aunque no siempre se comprueba. En nuestra vida cotidiana manejamos hipótesis continuamente, científicas, basadas en experiencias sesgadas, prejuicios, imaginarios colectivos, concepciones culturales...

Dos ejemplos

Hace unos años, cuando Power Point no había irrumpido todavía en el aula universitaria, mis colegas de asignatura y yo teníamos la hipótesis de que utilizar la proyección de imágenes en movimiento sobre elementos estadísticos, superando algunas limitaciones de la pizarra, incrementaría el interés, posiblemente la atención y, consecuentemente, mejoraría las calificaciones. Lo comprobamos. Un año impartimos nuestra docencia habitual de pizarra en tres de los grupos de clase. En los otros tres utilizamos la proyección de imágenes con un ordenador, explicando precisamente este tema: la decisión estadística mediante la prueba de significación de la hipótesis nula. Tuvimos cuidado de controlar variables extrañas, como utilizar el mismo tiempo, acudir a los mismos ejemplos, realizar los mismos ejercicios y aplicar el mismo examen de calificación final. Los resultados indicaban que: estábamos en lo cierto. Los grupos que asistieron a las clases con ordenador obtuvieron mejores notas por término medio.

No obstante, había una explicación alternativa. Tal vez no era el ordenador, sino la novedad. Andábamos por el año 1994. Con los medios que nos brindaba la universidad por entonces, tuve que trabajar mucho para un rendimiento dudoso. No teníamos dinero para el programa de ordenador que generaba las animaciones, así que programé un lenguaje de animación interpretado que llamé PIACI. Con él implementé nuestro guión. No teníamos técnicos especializados en estos asuntos. Pero encontré a un voluntario, el amigo de un amigo, especialista en hacer montajes caseros con una calidad muy aceptable para la época. Como pudimos, salió la animación. Y fue muy efectiva. Las cosas han cambiado bastante. Ahora creo que nosotros, los pioneros de aquel atrevimiento informático, somos los únicos que seguimos con la pizarra y no damos las clases con Power Point. Al entrar al aula encendemos las luces o abrimos las ventanas, para dar luz a espacios continuamente oscuros. Nuestra hipótesis ahora es que aquel efecto no se daría hoy. Ya no hay novedad. Es más, creemos que la gente está saturada de clases con diapositivas de Microsoft. Tal vez lo que seguimos haciendo, utilizar la pizarra (esta vez de rotuladores, y chiquitita, la verdad), sea hoy la novedad en un contexto de exceso tecnológico. Nos pasamos un poco, como hacemos siempre con la tecnología. Acaba utilizándonos. Y yo acabo de escribir aquí otra hipótesis.

La hipótesis del aumento de aprendizaje mediante el ordenador siguió un proceso científico y fue comprobada empíricamente. La hipótesis de que se acabó el efecto e incluso la de claudicar ante la tecnología tal vez estén sesgadas. No lo hemos comprobado, así que es posible que no estemos en lo cierto. En esto, y a diferencia del estudio anterior, nos estamos comportando de una forma habitual, puesto que las personas somos máquinas incansables que hipotetizan sin comprobación.

De lo que nos vamos a ocupar aquí es del primer esquema: hipótesis bien definidas que inspiran una investigación empírica, cuyos resultados permiten reforzarla o descartarla. No es necesario poner en marcha una recogida de datos tras cada hipótesis. Los datos nos rodean. E incluso muchos de ellos tienen formato estadístico y son suministrados por diversos organismos. Hay muchas personas cuya trayectoria profesional consiste en un continuo análisis de datos generados por otras.

Hipótesis de investigación, estadística y nula

Imagina que alguien afirma que la gente ya no fuma como antes. Es una hipótesis. Para comprobarla sería buena cosa obtener una muestra aleatoria de personas, proveniente de la población a la que se refiere la sentencia y ver qué pasa. Si vamos a poner en marcha un estudio para comprobar la sentencia, entonces la llamamos *hipótesis de investigación*. Para comprobarla es necesario contar con algún indicador que cuantifique el comportamiento de fumar. “Fumar como antes” es una expresión ambigua que admite muchas concreciones. Supongamos que se refiere a la cantidad de cigarrillos, con lo que redefinimos la sentencia así: “la gente ya no fuma el mismo número de cigarrillos que antes”. Pues bien, ya tenemos una medida cuantitativa, una variable: el número de cigarrillos fumados. No obstante habrá que acotarlo. Digamos que una semana es un tiempo suficiente para nuestros objetivos. Ahora necesitamos definir qué es eso de “antes”. ¿Tal vez el año pasado? Digamos que sí. Pero ¿cuánto fumaba la gente de esa población el año pasado? Pongamos que las estadísticas hablan de 30 cigarrillos a la semana.

Una hipótesis estadística consiste en un enunciado en el que se afirma un valor concreto para un estadístico. En nuestro ejemplo, hemos traducido la sentencia “la gente no fuma igual que antes” en la hipótesis estadística “la media de consumo hoy es diferente a 30 cigarrillos a la semana”. Habitualmente, ya que andamos en lenguajes simbólicos, expresamos las hipótesis de forma precisa con vocabulario matemático: si M es la media de consumo de cigarrillos que la gente fuma durante una semana el presente año, entonces la hipótesis estadística traducida directamente de la sentencia original afirma que $M \neq 30$.

La hipótesis nula (H_0) es una hipótesis estadística donde se afirma que el parámetro (la medida que nos interesa calculada en la población) no expresa cambio, efecto ni relación. En nuestro caso, H_0 afirmaría que no ha tenido lugar ningún cambio del año pasado al presente, que todo sigue igual, por lo que $H_0 \rightarrow M=30$.

Pensemos en la hipótesis de investigación “Existe relación entre el número de horas que un joven pasa escuchando música anglosajona y el interés que siente por la cultura inglesa o estadounidense”. El número de horas escuchando música se simboliza con H. El interés por la cultura inglesa o anglosajona se mide mediante un cuestionario que arroja puntuaciones cuasicuantitativas y se expresa con la variable C. Como veremos en otra unidad, un recurso estadístico para medir la relación entre dos variables cuantitativas es la correlación lineal simple de Pearson que se simboliza con la letra r. Si r_{HC}= 1, la relación entre ambas variables es máxima y positiva (a más horas, más interés). Si r_{HC}= -1, la relación es máxima y negativa (a más horas, menos interés). Y si r_{HC}= 0, no existe relación (horas e interés no tienen nada que ver entre sí). Pues bien, la hipótesis nula, al negar la relación entre ambas variables, establecerá $H_0 \rightarrow r_{HC}=0$.

Pero ¿por qué esta manía de la H_0 en contradecir nuestra ilusión depositada en encontrar que las cosas cambian o hay relaciones entre variables? La respuesta está en la justicia.

La lógica del proceso

La metáfora de la justicia

Sobre el papel, un Estado de Derecho se define más o menos como un lugar donde la justicia está garantizada al menos en términos de que todo el mundo es inocente mientras no se demuestre lo contrario. El esquema es muy conocido, especialmente a través de algunas de las miles de películas made in USA que vemos a lo largo de los años. Sintéticamente:

1. Se lleva a alguien a juicio porque se sospecha que ha cometido un delito.

2. El proceso comienza con una premisa: esa persona es inocente mientras no se demuestre lo contrario.
3. Se presentan evidencias o hechos que permiten dudar de la premisa inicial.
4. A la luz de lo que se deduce de esas evidencias o hechos, se toma una decisión: ¿se mantiene o se rechaza la hipótesis de inocencia?
5. En función de la decisión, tiene lugar una conclusión: si se rechaza la hipótesis de inocencia, se declara culpable al acusado. En caso contrario, el juicio termina manteniendo su inocencia.

El corazón lógico del proceso viene a ser: contamos con una hipótesis y con datos. Si los datos son compatibles con la hipótesis, esta se mantiene. Pero si no lo son, entonces damos más credibilidad a los datos que a las creencias, por lo que la hipótesis es rechazada.

En una prueba de significación de la hipótesis nula (PSHN) se lleva un proceso similar, que puede ser expuesto con los mismos puntos:

1. Se sospecha que existe algún cambio en un parámetro, alguna relación entre variables o alguna diferencia de algún tipo.
2. El proceso comienza con la hipótesis nula, representada con el símbolo H_0: el parámetro

sigue teniendo el mismo valor, no existe relación entre las variables o no hay diferencia entre los términos que se comparan, mientras no se demuestre lo contrario.

3. Se presentan evidencias o hechos que permiten plantearse la credibilidad de la H_0.
4. A la luz de lo que sededuce de esas evidencias o hechos, se toma una decisión: ¿se mantiene o se rechaza la H_0?
5. En función de la decisión, tiene lugar una conclusión que traduce el rechazo o el mantenimiento de la H_0 en los mismos términos con que se enunció el problema de investigación.

Sobre la compatibilidad entre datos e hipótesis

Un asunto relevante en el proceso es la *compatibilidad* entre los datos y la hipótesis. Constituye el meollo de la cuestión. Si son compatibles, se mantiene la hipótesis nula. Si son incompatibles, se rechaza. Pero ¿cómo medir la compatibilidad?

Hay que tener muy presente que la hipótesis se refiere siempre a la población, mientras que los datos provienen de la muestra. Sabemos que las muestras son como son, es decir, suministran estadísticos cuyos valores *rondan* los parámetros. Es perfectamente esperable y asumible cierta distancia entre un estadístico y un parámetro sin que esa falta de coincidencia exacta indique que la hipótesis no sea cierta. Imagina por ejemplo que se supone que la estatura media de una población es de 160 centímetros. Ahora seleccionamos una muestra al azar donde la altura media es de 157 centímetros. ¿Son compatibles la media muestral y la poblacional? Es obvio que 157 y 160 son dos cantidades diferentes. Pero pueden ser compatibles. Es decir, es posible esperar una muestra con media 157 que proviene de una población con media 160. Si la probabilidad de que eso ocurra es alta, entonces hay compatibilidad. Si es algo raro o poco esperable, entonces no hay compatibilidad. Con los datos que te he contado no es posible concluir. Si en esa población hay muy poca variación en la variable altura, es decir, si casi todo el mundo mide 160 centímetros o prácticamente todo el mundo se encuentra entre 158 y 162, por ejemplo, entonces es muy difícil obtener una muestra grande cuya media aritmética sea 157. Diremos entonces que esa muestra no proviene de esa población, pues datos e hipótesis no son compatibles. Decir “esta muestra no viene de esa población” es una forma de expresar “la población de donde procede esta muestra no tiene una altura media de valor 160”. Pero si la muestra es muy pequeña, de tal forma que fluctúa mucho, y la altura en la población varía mucho pues se encuentran personas de todo tipo en este sentido, entonces la diferencia entre 160 y 157 es despreciable, de tal forma que datos e hipótesis son compatibles.

Una persona alegre tiene días tristes. Es compatible. Pero si tiene muchos días tristes, ya no será creíble definirla como una persona alegre.

Así pues, la compatibilidad se mide en términos de en qué medida la distancia que existe entre lo observado (datos de la muestra) y lo esperado (hipótesis nula referida a la población) tiene suficiente entidad como para darle significado o "ser significativa". Por esa razón hablamos de *significación de la hipótesis nula*. Como lo que hacemos es ponerla a prueba mediante los datos, ya puedes entender de dónde viene la expresión del acrónimo PSHN.

Dos procedimientos para concretar la lógica

Lo esperable o compatible se puede identificar mediante dos caminos:

a) El intervalo donde se espera al parámetro es {9,7±0,5} = {9,2; 10,2}. Como 9 queda fuera del intervalo, es decir, como no es uno de los valores esperados, entonces se rechaza H_0.

2. Estimar qué debería ocurrir en la población, observando si eso es compatible con lo que afirma H_0. Este procedimiento consiste en construir una estimación por intervalo. Si el valor del parámetro que defiende H_0 se encuentra

dentro de los valores esperables, es decir, dentro del intervalo de confianza, entonces H_0 se mantiene. Si cae fuera, H_0 se rechaza. Imagina que se supone que el consumo de alcohol es de 9 litros equivalentes anuales por término medio. En la muestra se ha obtenido una media de 9,7 litros. Con una confianza del 95% hemos calculado un error de precisión de 0,5 litros. Las operaciones pueden ser:

a) La diferencia entre el estadístico y el parámetro es de 0,7 litros, mayor que la máxima distancia que cabe esperar (definición del error de precisión), es decir, 0,5 litros. Por ello, se rechaza H_0.

3. Calcular cuál es la probabilidad de encontrar resultados del tipo de los observados, suponiendo cierta H_0. Esta probabilidad, en sentido estricto, nunca será cero. Pero puede resultar tan pequeña que lleguemos a considerar que es muy raro o difícil que ocurra tal cosa, por lo que rechazamos H_0. Si me presentan a una persona que, según me dicen, es muy habladora, pero habla poco o nada cuando nos conocemos, es posible que yo le encontrara uno de esos días raros que tiene todo el mundo. No es algo *muy difícil*, por lo que sigo manteniendo la hipótesis de que se trata de alguien que habla

mucho. Pero si las cuatro o cinco ocasiones en que nos hemos cruzado resulta que prácticamente no ha dicho palabra, entonces dejaré de creer en su descripción de persona habladora. No es imposible que ocurra lo que describo en una persona catalogada como muy habladora. No es imposible, pero resulta tan difícil que finalmente creo más en mi experiencia (datos) que en la afirmación que se me ha comunicado.

En los dos siguientes apartados vamos a abordar estos dos procedimientos. El primero va a ser inmediato, puesto que ya sabemos construir intervalos de confianza. El segundo es necesario, ya que es el habitual en las publicaciones, informes y recursos informatizados de análisis.

En ambos casos establecemos una misma secuencia que nos lleva hasta la conclusión final en relación a qué hacemos con la hipótesis de investigación. El proceso se muestra esquematizado en el cuadro 1. Pensemos en el ejemplo del consumo de alcohol.

1) Enunciado de H_0.
2) Estudio empírico.
3) Decisión.
4) Conclusión.

Cuadro 1: esquema de una PSHN.

Pregunta de investigación: ¿Ha cambiado el consumo de alcohol en la población?

1. Enunciado de la hipótesis nula. $H_0 \rightarrow \mu = 9$.
2. Estudio empírico. Obtenemos la muestra. Se hacen los cálculos, con una media aritmética de 9,7 y un intervalo de $\{9,2; 10,2\}_{0,95}$.
3. Decisión. Como el valor del parámetro considerado por la hipótesis nula no se encuentra dentro del intervalo, entonces se rechaza H_0, recordando que lo hacemos utilizando un nivel de seguridad del 95%.
4. Conclusión. Sí, el consumo de alcohol en la población ha cambiado.

Antes de entrar en el siguiente apartado, observa lo escrito en el punto 3: “recordando que lo hacemos utilizando un nivel de seguridad del 95%”. Esta consideración es fundamental. El nivel de seguridad o de confianza utilizado para construir el intervalo ha sido del 95%. Sabemos que al modificar ese valor, la amplitud del intervalo varía. Si hubiéramos escogido una confianza del 99%, por ejemplo, tal vez el intervalo sería $\{8,9; 10,5\}_{0,99}$. En este segundo caso, el valor del parámetro que defiende la hipótesis nula se encontraría dentro del intervalo, sería compatible con los datos y, por tanto, la decisión habría sido mantener H_0, por lo que la conclusión sería “No, el consumo de alcohol sigue siendo el mismo”.

Este inciso no tiene por objetivo generar indefensión. Las decisiones no son arbitrarias. Recuerda que el nivel de confianza se deriva de la reflexión sobre las consecuencias que conlleva equivocarse. Si es una decisión intrascendente, podemos manejar una seguridad baja, construir un intervalo estrecho y concluir casi con cualquier tipo de cambio o relación. Pero si hay que practicar el principio de precaución, para no aventurar conclusiones sin un fundamento muy sólido, entonces exigiremos mucha seguridad, intervalos amplios y, con ello, dificultades para rechazar las hipótesis. Si el procedimiento no es más *automático* es porque sería peligroso que lo fuera. La reflexión, por muy incómoda que suela parecer, es aun más imprescindible.

PSHN mediante intervalos de confianza

Las personas somos pésimas máquinas aleatorias. Creemos que somos capaces de comportarnos como si realizáramos con éxito selecciones al azar. Pero no es así, por fortuna. Para comprobarlo, hemos seleccionado al azar (con un procedimiento no humano) a 50 personas. Le hemos pedido que escriban sobre un papel un número al azar comprendido entre 1 y 10. Si fuéramos capaces de comportarnos como máquinas aleatorias, los datos generados deberían responder a una distribución aleatoria más o menos uniforme, es decir, con una frecuencia similar para cada uno de los diez números enteros que hay entre 1 y 10. Dado que hay 50 personas, cada valor debería aparecer aproximadamente 5 veces y la media debería ser aproximadamente 5,5 (el centro exacto entre 1 y 10), entre otras consecuencias. No obstante, la experiencia demuestra que la gente suele irse hacia los extremos (1 y 10) y el centro (5), con alguna predilección por el valor 7. En un monográfico posterior vamos a estudiar directamente la diferencia que hay entre una repartición observada de frecuencias y una teórica. Ese objetivo excede este texto. Pero ahora nos podemos encargar de las otras dos inquietudes, ¿El valor Xi=5 es realmente más frecuente de lo que cabría esperar? ¿La media aritmética es 5,5?

Tienes a continuación los datos recogidos, la tabla de frecuencias (donde M representa a la media) y algunos estadísticos:

4	7	2	3	7	4	8	10	7	4
5	3	7	5	7	6	3	7	7	5
10	5	6	8	1	5	5	5	1	10
1	10	10	9	1	4	5	9	5	9
6	10	5	7	5	5	5	7	10	6

Vamos a abordar ambos objetivos con un nivel de seguridad del 95% y resolver la prueba de significación de la hipótesis nula acudiendo a un intervalo de confianza alrededor del estadístico medido en la muestra, observando si el valor teórico o esperado se encuentra en el intervalo.

Xi	fi	Xifi	(Xi-M)^2fi
1	4	4	64
2	1	1	16
3	3	6	27
4	4	12	16
5	13	52	13
6	4	20	0
7	9	54	9
8	2	14	8
9	3	24	27
10	7	63	112
Suma	50	250	292

n	Media	D.t.	p(Xi=5)
50	5,00	2,42	0,26

¿La media es 5,5?

Es obvio que la media no es 5,5 en la muestra obtenida. Pero la distancia entre 5,5 y 5 puede ser algo perfectamente asumible por las fluctuaciones propias del azar. O no. Vamos a verlo.

1. $H_0 \rightarrow \mu = 5,5$

2. Datos: con la información suministrada por la muestra construimos un intervalo de confianza con una seguridad del 95%. Podemos acudir a la distribución normal porque la muestra es grande

(n = 50 ≥ 30), por lo que esa seguridad se traduce a una distancia estandarizada de valor 1,96. Con ello:

$$e_p = Z \frac{S}{\sqrt{n-1}} = 1{,}96\ \frac{2{,}42}{\sqrt{50-1}} = 0{,}68$$

$$\mu \in \{\bar{X} \pm e_p\}_{seg} \quad \Rightarrow \quad \mu \in \{5 \pm 0{,}68\}_{0{,}95} = \{\ 4{,}32\ ;\ 5$$

3. Decisión: como puede observarse, el valor del parámetro que establece la hipótesis nula se encuentra dentro del intervalo. Es, por ello, uno de los valores esperables. Luego, se mantiene H_0 manejando un nivel de seguridad del 95%.

4. Conclusión: sí, la media poblacional de números escogidos al azar en el intervalo 1 a 10 es 5.

El porcentaje de datos con el valor 5 ¿es del 10%?

En la muestra desde luego que no, pues observamos un 26%. Pero vamos a ver qué ocurre respondiendo convenientemente a esa pregunta. Antes hemos de comprobar si podemos suponer que la distribución muestral de proporciones es normal. Y así es, pues tanto nπ (50*0,10=5) como n(1-π) (50*0,9=45) cumplen la condición de ser iguales o superiores a 5. Observa que hemos utilizado π y no *p* muestral. La condición se expresa originalmente para π, no para *p*. No obstante, cuando hacemos una estimación por intervalo es porque queremos conocer precisamente π y, al no tenerlo, acudimos a *p*. No obstante, en una PSHN sí que contamos con π, la que establece la H_0, la que es cierta mientras no se demuestre lo contrario. Esta circunstancia tiene también efecto en el cálculo del error tipo. Recuerda que su expresión de cálculo está en función de π, pero que por las mismas razones ya expuestas, acudimos a *p*. No obstante, ya que tenemos un π (mientras no se demuestre lo contrario), hemos de acogernos a él. Vamos pues con el proceso:

1. $H_0 \rightarrow \ \pi = 0{,}10$

2. Datos: con la información suministrada por la muestra construimos un intervalo de confianza con una seguridad del 95%, que se traduce a una distancia estandarizada de valor 1,96. Con ello:

$$e_p = Z\sqrt{\frac{\pi\,(1-\pi)}{n}} = 1{,}96\sqrt{\frac{0{,}10 * 0{,}90}{50}} = 0{,}083$$

$$\pi \in \{\, p \pm e_p \,\}_{seg} \qquad \Rightarrow \qquad \pi \in \{0{,}26 \pm 0{,}083\,\}_{0{,}95} = \{\, 0{,}17$$

3. Decisión: como puede observarse, el valor del parámetro que establece la hipótesis nula no se encuentra dentro del intervalo. Luego, se rechaza H_0 manejando un nivel de seguridad del 95%.

4. Conclusión: no, el porcentaje poblacional de números escogidos al azar en el intervalo 1 a 10 con el valor 5 no es del 10%.

Antes de seguir es importante tener claro algo. En el primer ejemplo hemos mantenido la hipótesis nula concluyendo con un *sí*. En el segundo, hemos rechazado concluyendo con un *no*. Es una casualidad, no saques conclusiones precipitadas. Se puede concluir con un *no* manteniendo o con un *sí* rechazando. El meollo de la cuestión es no solo qué dice la hipótesis nula y qué dicen los datos, sino también cómo está enunciada la pregunta que expresa la hipótesis de investigación. Como has observado, la conclusión es una traducción de la decisión, adaptada como un guante a la pregunta que queremos responder con la PSHN. Voy a variar las dos preguntas de estos dos ejemplos, respondiendo correctamente según los datos. Verás que ya no coinciden el *sí* y el *no*.

Pregunta	Respuesta
¿La media es 5,5?	sí, la media poblacional de números escogidos al azar en el intervalo 1 a 10 es 5.
¿La media ha dejado de ser 5?	no, la media poblacional de números escogidos al azar en el intervalo 1 a 10 sigue siendo 5.
El porcentaje de datos con el valor 5 ¿es del 10%?	no, el porcentaje poblacional de números escogidos al azar en el intervalo 1 a 10 con el valor 5 no es del 10%.
El porcentaje de datos con el valor 5 ¿es distinto del 10%?	sí, el porcentaje poblacional de números escogidos al azar en el intervalo 1 a 10 con el valor 5 es distinto del 10%.

PSHN mediante el cálculo de probabilidades (y distancias)

Decidir es una cuestión de nivel

Aunque utilizar intervalos de confianza para responder a PSHN es muy comprensible y enlaza perfectamente con lo que ya sabíamos, no es lo habitual. Lo frecuente es acudir a valores de probabilidad. La razón principal es que permite más agilidad, es más cómodo para correr. Los programas de ordenador te regalan listados de resultados de análisis que incluyen valores de probabilidad. Y estos se leen y supuestamente se comprenden con mucha mayor velocidad y con menos extensión que si todas las decisiones vinieran acompañadas por intervalos de confianza. Ya sabes que vives en una época donde todo el mundo tiene mucha prisa. En ciencia ocurre lo mismo.

Observa la siguiente tabla. Hemos pedido a un famoso programa de ordenador dedicado al análisis de los datos (SPSS) que lleve a cabo un conjunto de correlaciones entre pares de variables. Como veremos en otra unidad y hemos adelantado ya más atrás, la correlación pretende medir el grado de relación entre dos variables cuantitativas en el intervalo (-1, +1). Conforme más cerca se encuentre el valor de correlación de cualquiera de los extremos (-1 o +1), mayor es la relación entre ambas variables. Cada combinación de la tabla es una PSHN expresada velozmente en valores de probabilidad. Ahora mismo no sabemos interpretarla. Pero esto vamos a solucionarlo en breve. Lo que interesa ahora es conocer que un hábito muy extendido es pedir a un programa de ordenador que lleve a cabo una gran cantidad de PSHN al mismo tiempo, resueltas cada una de ellas mediante una comparación de probabilidades y no mediante intervalos de confianza. Hacer las cosas rápido no es bueno ni malo en sí mismo. Lo malo es no saber lo que se está haciendo.

	edad	altura	peso	mates
relación de Pearson	1	,395	,574(**)	-,31
(bilateral)		,056	,003	,14
	24	24	24	2
relación de Pearson	,395	1	,701(**)	-,33
(bilateral)	,056		,000	,10
	24	24	24	2
relación de Pearson	,574(**)	,701(**)	1	-,16
(bilateral)	,003	,000		,44
	24	24	24	2
relación de Pearson	-,311	-,337	-,165	
(bilateral)	,140	,107	,442	
	24	24	24	2
relación de Pearson	-,168	,157	,128	,409(*
(bilateral)	,432	,464	,550	,04
	24	24	24	2

** La correlación es significativa al nivel 0,01 (bilateral).

* La correlación es significante al nivel 0,05 (bilateral).

Vamos a llegar a las probabilidades desde lo que conocemos, los intervalos de confianza. Una forma de ver cómo se construye un intervalo de estimación es observar las sucesivas traducciones que tienen lugar: de un valor de probabilidad (el nivel de seguridad o de confianza) a una distancia estandarizada, de esta a un valor para el error de precisión, y de este a un intervalo. El error de precisión, como sabes, expresa la máxima distancia que cabría esperar por azar entre el estadístico y el parámetro. Pues bien, si contamos con un resultado muestral (estadístico) y un supuesto valor poblacional según H_0 (el parámetro), podemos invertir el proceso: pasar de la distancia entre ambos a una distancia estandarizada y de esta a un valor de probabilidad. La tabla 1 expresa esta idea.

Nivel	Observado	Teórico
Probabilidad	(grado de sign.)	(nivel de sign.
	p (6)	α (1)
	↑	↓
Distancia estandarizada	Z_{obs} (5)	Z_{seg} (2)
	↑	↓
Puntuación directa	diferencia (4)	e_p (3)

Tabla 1. Niveles de comparación.

La tabla 1 empareja los dos términos que podemos comparar en cada caso. Si la diferencia observada (casilla 4) es superior a la esperada (casilla 3), se rechaza H_0; en caso contrario se mantiene. Lo mismo ocurre en términos estandarizados (casillas 5 y 2, respectivamente). Mientras que ocurre lo contrario con las probabilidades: si la probabilidad de errar al rechazar H_0 o riesgo calculado (casilla 6) es superior al máximo riesgo que desearíamos asumir (casilla 1), H_0 se mantiene; en caso contrario se rechaza. Estas lógicas son las que vamos a desgranar a continuación.

Para comprender esta tabla, vamos a partir de un ejemplo ya conocido: el de la media aritmética de los números supuestamente aleatorios que ha suministrado una muestra de 50 personas.

Comparando distancias en puntuaciones directas

Lo teórico o esperado, es decir, la sentencia contenida en H_0, indica que la media aritmética debería ser 5,5. Lo observado o medido en la muestra es 5. La diferencia entre ambos es 0,5. Pues bien, este valor (que se corresponde con la casilla 4 de la tabla 1) tiene una interpretación difícil. No sabemos qué hacer con ese 0,5. Sabemos que no es 0, es decir que existe alguna diferencia entre lo observado y lo esperado. Pero tal vez esa diferencia no tenga ningún significado, no sea una *diferencia significativa*. El valor 5 en la media aritmética es uno de los muchos que pueden suministrar las muestras aleatorias. Tal vez la diferencia 0,5 sea achacable a las fluctuaciones propias del azar.

Tal vez tenga suficiente entidad como para rechazar el azar como explicación.

Para solucionar ese problema, iniciamos un proceso que comienza en la casilla 1: decidimos con qué seguridad queremos llevar a cabo la decisión. En las PSHN, no suele pensarse tan en positivo (seguridad o confianza) sino en términos de error. Si la seguridad o confianza es del 95%, entonces la desconfianza, inseguridad o probabilidad de errar es 5%. Este 5% recibe también otros nombres similares, como riesgo de equivocación (así lo hemos nombrado al estudiar los intervalos de estimación estadística) o *nivel de significación* (así es como suele conocerse cuando se aborda la decisión estadística con PSHN), que remarca nuestro objetivo: estamos evaluando en qué medida es significativa la diferencia encontrada. Se simboliza con la letra griega alfa: α (otras grafías: ɑ, α, α, α...).

Pues bien, lo que hemos hecho en los ejemplos anteriores donde la PSHN se ha basado en un intervalo de confianza ha sido traducir ese nivel de significación (o su complementario, la confianza) en una distancia estandarizada. En nuestro caso es 1,96. Ya nos encontramos en la casilla 2. Una vez con ese valor de 1,96 en la mano, hemos calculado el error de precisión, que dio como resultado 0,68. Ya estamos en la casilla 3. Y ahora ya podemos comparar. Nos encontramos en el nivel de las puntuaciones directas tanto en el plano de lo teórico (no esperamos una distancia superior a

0,68) como de lo observado (la distancia ha sido de 0,5). En el ejemplo del apartado anterior, no finalizamos aquí sino que dimos un paso más construyendo el intervalo de confianza y observando que el parámetro se encontraba dentro, por lo que datos e H_0 eran compatibles y mantuvimos H_0. Pero este último paso, de restar y sumar al estadístico el valor del error de precisión, no era necesario. Basta con comparar la distancia observada (0,5) con la máxima que cabe esperar por azar considerando una confianza del 95% (0,68). Dado que lo observado (0,5) está dentro de lo esperado (no más de 0,68), entonces mantenemos H_0.

Comparando distancias estandarizadas

En lugar de comparar al nivel de las puntuaciones directas, podríamos hacerlo con puntuaciones estandarizadas. La teórica ya la conocemos, Z_{seg}=1,96. Para establecer la comparación necesitamos estandarizar la distancia observada (0,5). Como sabes, estandarizar es traducir una puntuación directa expresando el número de desviaciones tipo que la separan de su media. Pues bien, 0,5 ya es distancia de la puntuación observada (media de valor 5) a la media de la distribución muestral de medias, valor esperado de la media o media de la población (5,5). Ahora nos resta expresarla en número de desviaciones tipo.

$$Z_{obs} = \frac{|\bar{X}_i - \mu|}{\sigma_{\bar{X}}} \simeq \frac{|\bar{X}_i - \mu|}{\frac{S}{\sqrt{n-1}}} = \frac{|5 - 5{,}5|}{\frac{2{,}42}{\sqrt{50-1}}} = 1{,}45$$

En términos estandarizados, esperamos una distancia de valor Z_{seg}=1,96 (casilla 2) como mucho. Lo observado es Z_{obs}=1,45 (casilla 5). Por tanto, lo observado es compatible con lo esperado y mantenemos H_0. Hemos establecido la comparación en el nivel de las puntuaciones estandarizadas.

El recurso de la distancia estandarizada Z es el habitual cuando las distribuciones muestrales son normales. En muchas situaciones de investigación no se puede suponer una distribución normal. Es más, muchas PSHN tienen su propio recurso similar a Z pero partiendo de otra distribución de referencia, con su propio nombre y símbolo. NO obstante, la lógica es la misma que estamos conociendo para el caso de las Z. Así, por ejemplo, cuando queremos concluir si un conjunto de frecuencias (como, por ejemplo, el número de votos que ha recibido cada uno de entre un listado de partidos políticos) difiere o no de lo que cabría esperar, utilizamos un estadístico que se llama chi cuadrado o ji cuadrado y se simboliza con χ^2. Pues bien, la decisión en la PSHN se lleva a cabo del mismo modo que estamos viendo en el caso de las Z: si χ^2_{obs}

(calculado con los datos de la muestra) es superior a lo que cabría esperar por azar (χ^2_{seg}, buscado en una tabla, a partir del nivel de seguridad que deseamos tener en la decisión), entonces se rechaza la hipótesis nula. Si nos encontramos en una situación donde queremos concluir si un tratamiento ha reducido el nivel de ansiedad de un grupo de personas, comparando su ansiedad media antes y después del tratamiento, suele acudirse a una PSHN basada en un estadístico que se denomina *t de Student*. Del mismo modo, si t_{obs} (calculada con los datos de la muestra) es mayor de la que cabría esperar por mero azar según el nivel de seguridad que estamos manejando (t_{seg}), entonces se rechaza la hipótesis nula.

Comparando probabilidades

Aun podemos dar un paso más allá. En lugar de traducir el nivel de significación del 5% a una distancia estandarizada máxima de valor 1,96, lo que podemos hacer es traducir la distancia estandarizada observada (1,45) a un valor de probabilidad que, para distinguirlo del anterior, denominaremos *grado de significación*. El nombre está muy bien escogido. Si bien el *nivel* es estático (lo hemos establecido mediante reflexión o tradición antes de contar con los datos), el *grado* es dinámico, es un cálculo que deriva de las características específicas de la muestra. El nivel es como la altura en que se coloca el listón en un salto de altura de una prueba de atletismo. El grado es como lo que puede saltar realmente el deportista. Si salta por encima del nivel, superó la prueba. Si salta por debajo, tropieza con el listón y la prueba no dio el resultado deseado. En PSHN, el deportista es H_0. Si pasa la prueba, se mantiene, y esto es porque el grado fue superior al nivel.

El grado de significación es la probabilidad de encontrar resultados como los observados (sean exactamente esos o aun más raros) suponiendo cierta la hipótesis nula. Si esa probabilidad es muy alta, es que esos resultados no son tan raros y seguimos suponiendo cierta H_0, es decir, la mantenemos. Pero si esa probabilidad es muy baja, entonces es muy difícil asumir que H_0 sea cierta. ¿Qué cataloga al valor del grado de significación como alto o bajo?: el nivel de significación. Recuerda que es la medida de desconfianza, inseguridad o riesgo de error. ¿Error de qué? Si tenemos una confianza del 95% de que el parámetro esté dentro del intervalo, es que tenemos una inseguridad de que esté fuera del 5%. Si el parámetro cae fuera del intervalo de confianza no significa que H_0 sea falsa *necesariamente,* sino muy difícil de ocurrir, poco probable. La probabilidad de equivocarnos al tomar esa decisión de rechazo es 5%.

Pues bien, en lugar de traducir 5% de nivel de significacación (o 95% de nivel de confianza) en la

Z_{seg}=1,96, lo que hacemos ahora es traducir la Z_{obs}=1,45 a un grado de significación por el camino inverso. Al buscar Z=1,45 en la tabla de la curva normal estandarizada, esta distancia hasta la media está asociada con un área de 0,426 (fila 1,40 y columna 0,05). Es la superficie de la curva normal comprendida entre Z=1,45 y la media. Como nos vale tanto una distancia por exceso como por defecto, el área de confianza es 0,426x2=0,852. El grado de significación será 1-0,852=0,148.

Así pues, la probabilidad de errar al rechazar la hipótesis de que la media es 5,5 es del 14,8%. Esto es muy elevado puesto que el umbral máximo es 5%. Luego, no corremos este riesgo y mantenemos H_0.

Observa que, como no podía ser de otro modo, hemos mantenido la hipótesis nula utilizando el intervalo de confianza, la comparación de distancias en puntuaciones directas, la comparación de distancias estandarizadas y la comparación de probabilidades.

Observa ahora la tabla de correlaciones entre edad, altura, peso, mates y lengua. Ha surgido de preguntar a un conjunto de estudiantes de instituto por sus valores en esas cinco variables. Hemos utilizado el programa SPSS y nos ha respondido con esa tabla. Muchas personas no reflexionan sobre el proceso, ni realizan un estudio descriptivo previo, ni se detienen a interrogarse acerca del significado que tiene cada valor de correlación, sino que miran únicamente el valor de *significación estadística*, que en la tabla aparece como *Sig*. Por ejemplo, entre las variables *edad* y *altura* existe un valor de correlación de cuantía 0,395. Esto significa que ambas variables tienen una relación positiva (a más edad, más altura), puesto que ese valor se encuentra entre 0 y 1. Si nos planteamos si 0,395 es estadísticamente significativo, es decir, si ese valor se aleja suficientemente de 0 como para suponer que en la población no es 0 (0 representa “no relación entre ambas variables), el grado de significación asociado sería 0,056. Si estamos manejando un nivel de significación de 0,05 y dado que 0,056 > 0,05 entonces deberíamos tomar la decisión de mantener la hipótesis nula y concluir que ambas variables no están relacionadas, es decir, que no existe relación entre la edad y la altura. SPSS lo pone todavía más sencillo: si el grado de significación es inferior a 0,05 entonces lo marca con un asterisco (como ocurre en la relación entre mates y lengua), mientras que utiliza dos asteriscos para señalar un grado de significación inferior a 0,01 (como ocurre en la relación entre edad y peso). Los asteriscos permiten tomar decisiones con todavía mayor velocidad. Traigo aquí este resultado con el ánimo de llamar la atención: no es cuestión de ir más rápido sino de saber qué estamos haciendo, sea con más o con menos velocidad. Hay muchos inconvenientes asociados a este hábito veloz. Uno de ellos es que las relaciones pueden ser de muchos tipos y si no las estudiamos antes gráficamente, podemos estar haciendo tonterías yendo directamente a una PSHN. Otro inconveniente es el efecto de lo que se denomina “comparaciones múltiples”. ¿Recuerdas el dicho “tanto va el cántaro a la fuente que al final se rompe”? Si se hacen muchas PSHN de forma simultánea, hay que tener en cuenta que podemos estar obteniendo resultados estadísticamente significativos por pura casualidad. Ocurre también que si el tamaño de la muestra es muy grande, resulta difícil rechazar una hipótesis nula, aunque se ponga a prueba un valor muestral muy pequeño, incluso ridículo. De este modo, estamos concluyendo que existe significación cuando una breve reflexión sobre la cuantía del estadístico nos está diciendo que ahí no pasa nada. No te preocupes ahora por estas cuestiones. Las abordaremos en una unidad posterior. De momento, intenta comprender lo que estamos haciendo cuando se toman decisiones con la PSHN, sea con puntuaciones directas, con puntuaciones estandarizadas o con probabilidades. He sacado este asunto aquí porque es aquí donde estamos teniendo el primer contacto con las PSHN. Es bueno saber desde el principio que hay hábitos ya muy asentados que sería mejor no reproducir automáticamente por tu parte.

Vamos a por el otro ejemplo

Abordemos ahora el caso del porcentaje de personas que han escogido el valor 5 como número al azar. Encontramos que era 26%, mientras que cabría esperar un 10%. Vamos a tomar la misma decisión de rechazar H_0 mediante los tres procedimientos.

- Comparando distancias

Ya hemos hecho parte del trabajo en el apartado sobre PSHN mediante intervalo de confianza. Encontramos que el error de precisión, es decir la máxima distancia entre estadístico y parámetro que cabría esperar era de 0,083 u 8,3%. Pero la distancia observada es de 16%. Como se escapa de lo que cabría esperar por mero azar, consideramos que se trata de una distancia significativa y se rechaza H_0.

- Comparando distancias estandarizadas

La distancia estandarizada teórica es 1,96. Para estandarizar 16% o 0,16 hemos de acudir a la expresión ya conocida:

$$Z_{obs} = \frac{|p - \pi|}{\sigma_p} \simeq \frac{|p - \pi|}{\sqrt{\dfrac{\pi\,(1 - \pi)}{n}}} = \frac{|0{,}26 - 0{,}10|}{\sqrt{\dfrac{0{,}10 \cdot 0{,}90}{50}}} = 3{,}77$$

Dado que Z_{obs}=3,77 es sensiblemente mayor que Z_{seg}=1,96, es decir dado que lo encontrado se aleja sensiblemente de lo esperado si H_0 fuera cierta, se rechaza H_0.

– Comparando probabilidades

Sabemos que α=0,05 o 5%. Necesitamos el valor del grado de significación. Para ello, consultamos la tabla de la curva normal estandarizada para identificar el área, proporción o superficie de la curva normal que se aleja de la media en no menos de 3,77 desviaciones tipo. Y no tenemos información porque se nos termina la tabla. Sabemos que por encima de Z=3,3 lo que tenemos es toda la curva con una preción de tres decimales. Luego, una Z=3,77 es altamente improbable, lo suficiente como para que el grado de significación sea 0,000 con tres decimales (estrictamente no es cero, con más decimales de precisión obtenemos 0,000163). Como 0,000 es claramente inferior al máximo riesgo que asumiríamos para rechazar H_0, la rechazamos con claridad.

Un apunte final sobre símbolos

El grado de significación es el valor que suele acompañar a tablas de resultados, artículos científicos y salidas de programas estadísticos de ordenador. En cada uno de estos soportes puede utilizarse un nombre o símbolo diferente: *p*, *p-value*, *valor p*, *sign.*, *g.s.*. Es importante identificar el recurso concreto que se utiliza en cada caso. Es fácil identificar porque es calculado, no decidido. Si fuera decidido, estaríamos hablando del nivel de significación.

¿Cuántas colas tiene la cosa?

Cuando una persona o un grupo vive que una situación es negativa, lo más probable es que desee que cambie. Si un genio de la lámpara maravillosa cayera cerca y concediera el deseo tal vez resultara peor el remedio que la enfermedad: que algo cambie no significa que mejore. Es una obviedad que, como muchas obviedades, no es raro que se nos pase por alto. Algunas personas que están mal, por ejemplo, buscan la esperanza en opciones todavía peores. Ocurre en procesos de inmigración, en el comportamiento de muchos votantes, en roturas o inicios de parejas, etc. Es cotidiano.

El objetivo de este apartado no es expresamente adentrarnos en aspectos fundamentales de la vida. Más bien nos conformaremos con algo de estadística. En concreto nos preocupa trabajar con hipótesis que no se conformen con el simple cambio sino que consideren también el sentido del cambio. En la siguiente tabla contamos con dos columnas con preguntas de investigación muy parecidas pero diferentes. La columna de la izquierda plantea dudas que se responden con una hipótesis del tipo que hemos conocido: que no reconoce el sentido del cambio sino solo el cambio, es decir, para la que tanto vale un exceso como un defecto, un alejamiento por la derecha como por la izquierda. Es el caso de las pruebas denominadas “de dos colas”, denominación que apunta a esos dos sentidos o extremos aceptables. La columna de la derecha ejemplifica inquietudes de una sola cola, es decir, donde el sentido del cambio es fundamental: o debe ser una cantidad superior, o debe serla inferior, pero no cualquiera de ambas indistintamente.

s	Una cola
o el consumo de alcohol respecto dos años?	¿Se bebe más alcohol ahora qu años?
écnica de intervención ¿modifica el siedad?	La nueva técnica de intervenció la ansiedad?
y mujeres difieren en su gusto por s del Oeste?	¿Gustan más las películas del C hombres que a las mujeres?
to la cadena de televisión C3 que	¿C3 es una cadena de televisió seguida que H4?
inario ¿la gente ha modificado su	¿El seminario ha rebajado los n

s	Una cola
nofobia?	xenofobia?
ncia hacia el otro es diferente en que en inmigrantes?	¿Son los autóctonos menos tole el otro que los inmigrantes?

Tabla 2. Ejemplos de una y dos colas.

Las PSHN que hemos ejemplificado en este documento hasta el apartado anterior se corresponden con una prueba de dos colas. Esto es lo más inmediato cuando se parte de una estimación por intervalo. Ten en cuenta que en un intervalo de confianza consideramos todo el área que se encuentra relativamente cercana a la media o valor esperado, rechazando los extremos, tanto el de los valores más altos como el de los valores más bajos. Un intervalo, por defecto centrado, no es la mejor opción para una PSHN de una sola cola. Las comparaciones mediante distancias (sean directas o estandarizadas) o de probabilidades son recursos más naturales o menos forzados que acudir a un intervalo.

La lógica de una prueba de una cola, frente a la de dos, se puede expresar en dos puntos:

1. Al observar el resultado empírico podemos saber si continuamos con el proceso o no. Si estamos en una prueba de una cola a la derecha, es

decir, si estamos probando que existe una diferencia positiva o por arriba, pero observamos en la muestra que la diferencia es negativa o por debajo, entonces carece de sentido probar nada. Ya se ve claro que hay que mantener H_0. Por ejemplo, si creemos que las mujeres fuman más que los hombres y ya en la muestra se ve que son los hombres los que fuman más que las mujeres, no importa si esta diferencia es o no significativa, lo relevante es que se encuentra en el sentido opuesto al que nos preocupa, por lo que no pasamos a la fase probabilística de PSHN y ya concluimos que no podemos afirmar que las mujeres fumen más que los hombres.

2. Si el sentido de la diferencia es acorde con nuestras sospechas, entonces nos ocupa valorar si esa diferencia es significativa o tiene suficiente entidad como para mantener que también es defendible su ocurrencia en la población y no solo en la muestra. En ese sentido, realizamos cualquiera de las comparaciones de nivel que hemos visto, pero con una salvedad: la traducción del nivel de significación o riesgo de equivocación debe actualizarse. Así, por ejemplo, un riesgo de $\alpha=5\%$ se corresponde con una $Z_{seg}=1,96$ solo en el caso de dos colas, puesto que el 5% se reparte en ambas a un

2,5% por cola. Por esta razón, en muchas publicaciones puedes encontrar Z_{seg} expresada como $Z_{\alpha/2}$. Esto deja un área centrada del 95% y, por tanto, una área hasta la media de 47,5%. Pero si el α=5% se encuentra íntegro en un extremo, el área hasta la media es del 45%, por lo que Z_{seg}, que puede encontrarse en los textos como Z_{α}, ya no es 1,96 sino que Z_{seg}=1,645. Es la misma distancia estandarizada que se corresponde con un riesgo de dos colas de α=10%, puesto que en este caso dejamos un 5% a cada extremo o en cada cola. Las figuras 1a, 1b y 1c expresan esto mismo pero gráficamente. Todas se han generado para un riesgo del 5%. En 1a se observa una prueba de una cola donde el área o probabilidad de rechazo se encuentra a la izquierda. En 1b es una prueba de dos colas (el riesgo está repartido por la derecha y por la izquierda). Mientras que en 1c, los valores que llevan a rechazar han de ser aquellos que exceden en la zona del 5% superior.

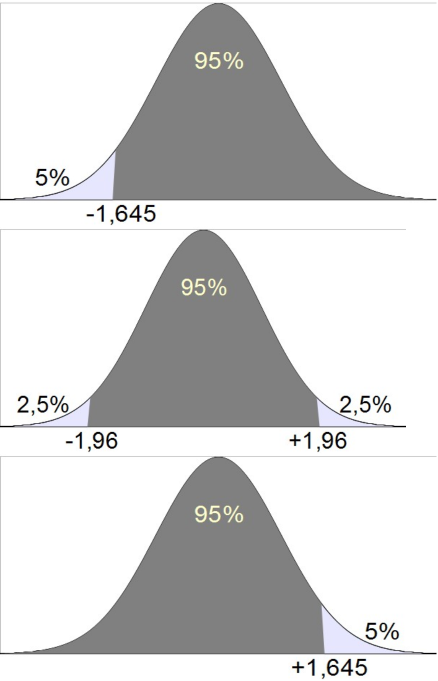

Figura 1. a) cola izquierda; b) dos colas; c) cola derecha.

Veamos un ejemplo. Se considera que la ansiedad media de una población es de 3,4 puntos en una escala de 1 a 7. Pero algunos indicios parecen sospechar que la población ha reducido su ansiedad. Para comprobarlo, preguntamos a una muestra de n=66 personas de esa población, obteniendo los siguientes resultados:

2	3	1	7	2	5	6	5	6	7
3	5	1	2	6	1	5	7	3	3
5	1	3	5	3	2	5	7	4	6
1	2	7	7	1	3	7	5	5	4
1	1	3	1	7	2	2	5	6	2
5	6	7	6	5	6	4	5	4	1

La tabla de frecuencias y algunos estadísticos son:

Xi	fi	Xifi	(Xi-M)^2fi
1	10	10	90
2	9	18	36
3	10	30	10
4	5	20	0
5	15	75	15
6	8	48	32
7	9	63	81
Suma	66	264	264

n	Media	D.t.
66	4	2

Según podemos ver, la media aritmética de la muestra tiene el valor 4. Es superior a lo que cabría esperar, mientras que la sospecha es de que la ansiedad había disminuido. No hace falta seguir con el proceso, puesto que ya en la muestra los datos contradicen la hipótesis de investigación.

No obstante, imaginemos que la hipótesis, también de una cola, es que la ansiedad ha aumentado. En ese caso, tras recoger los datos y analizarlos, comprobamos que la muestra suministra resultados acordes con la hipótesis de investigación. Al menos al nivel de la muestra se cumple con lo que se espera. Lo que procede es comprobar si podemos seguir asumiéndolo en la población. Es un problema típico para la PSHN. que vamos a resolver mediante los tres procedimientos que hemos aprendido comenzando, por ejemplo, con la comparación de distancias directas. Vamos a considerar un nivel de significación del 3%.

Como se trata de una muestra grande ($n = 66 \geq 30$), podemos suponer que proviene de una distribución muestral normal. Por otro lado, al tratarse de una prueba de una cola, el área extrema del 3% se encuentra únicamente en una parte de la curva. La distancia estandarizada que acota esa superficie superior es la misma que acotaría un 6% de ambos extremos, un 94% central o un área hasta la media del 47%. Con todas estas formas de concebir esa misma distancia podemos consultar diversas tablas. Todas ellas me indicarán que Z_{seg} o $Z_{\alpha} = 1{,}88$. Considerando los resultados muestrales:

$$e_p = Z \frac{S}{\sqrt{n-1}} = 1{,}88 \; \frac{2}{\sqrt{66-1}} = 0{,}47$$

1. $H_0 \rightarrow \mu \geq 3{,}4$
2. Resultados: $\bar{X} = 4$, separada del valor esperado en 0,60 unidades. Dado que esperamos que la máxima distancia aleatoria sea de 0,47 unidades, lo encontrado se escapa de lo esperado. Luego,
3. Decisión: se rechaza H_0 utilizando un nivel de significación del 3%.
4. Conclusión: la media de ansiedad ha aumentado.

Si realizamos la PSHN mediante distancias estandarizadas, ya sabemos que Z_{seg} o Z_α = 1,88. Lo que hacemos ahora es traducir la distancia observada en estandarizada:

$$Z_{obs} = \frac{|\bar{X}_i - \mu|}{\sigma_{\bar{X}}} \simeq \frac{|\bar{X}_i - \mu|}{\dfrac{S}{\sqrt{n-1}}} = \frac{|4 - 3{,}4|}{\dfrac{2}{\sqrt{66-1}}} = 2{,}42$$

Con ello:

1. $H_0 \rightarrow \mu \geq 3{,}4$
2. Resultados: $Z_{obs} = 2{,}42$
3. Decisión: Como $Z_{obs} > Z_{seg} = 1{,}88$, se rechaza la H_0 utilizando α=0,03.

4. Conclusión: La media de ansiedad ha aumentado.

Si llevamos a cabo la PSHN mediante probabilidades, ya contamos con una de ellas, el nivel de significación, cuyo valor es α=0,03. Para calcular el grado de significación hay que acudir a las tablas con nuestra Z_{obs} = 2,42, teniendo en cuenta que estamos manejando una prueba de una cola. Según la tabla de áreas hasta la media, una distancia estandarizada de valor 1,88 se corresponde con un área hasta la media de 0,492. Como nos interesa solo una zona, cola o lado de rechazo, todo el otro lado de la curva no interesa, por lo que g.s.= 0,5 – 0,492 = 0,008. Luego:

1. $H_0 \rightarrow \ \mu \geq 3{,}4$
2. Resultados: g.s. = 0,008
3. Decisión: Como g.s. < α = 0,03 entonces se rechaza H_0
4. Conclusión: La media de ansiedad ha aumentado.