

Citar como:

Frías-Navarro, D. (Diciembre, 2008). *Evaluación empírica de las hipótesis*. Universidad de Valencia. <http://www.uv.es/friasnav>

Evaluación empírica de las hipótesis

Frías-Navarro, Dolores
Universidad de Valencia
España

En el proceso del diseño de investigación un elemento clave es la *evaluación empírica de las hipótesis*. En los últimos años la ‘evaluación empírica de las hipótesis’ siempre va acompañada de los términos ‘basada en la evidencia’. El término ‘evidencia’ (en realidad la traducción al castellano de *evidence* es prueba) está de moda en los títulos de los trabajos de investigación. Podemos decir que el término se ha convertido en un icono, en un lema que todo investigador que se precie debe seguir. Porque ¿quién podría estar en contra de basar las decisiones en la evidencia o en pruebas? Lo que sí está claro es que existen diferentes grados de evidencia vinculados con el tipo de metodología utilizada en el diseño de investigación donde el ensayo aleatorizado y controlado (metodología experimental) ocupa la primera posición como la regla de oro de los diseños, aportando las pruebas de mayor calidad y asegurando con ello, en gran medida, la validez de las relaciones encontradas entre las variables.

Inferencia estadística

Una afirmación empírica es posiblemente verdadera y posiblemente falsa. Para poder establecer su veracidad procedemos a la recolección de datos. En la mayor parte de las ocasiones el proceso de evaluación empírica de las hipótesis se realiza mediante un modelo de decisión probabilística a través de la ejecución de una prueba o test de significación estadística. La forma más común consiste en ejecutar la *prueba de significación de la hipótesis nula* (*Null Hypothesis Significance Testing*, NHST). Este

procedimiento tiene una larga historia y una gran tradición en su aplicación por parte de los investigadores pero también ha recibido muchas críticas (véase una revisión en Nickerson, 2000 y Pascual, Frías y García, 2000) que han provocado cambios importantes en las normas de publicación científica desde finales del siglo XX dando lugar a la denominada “reforma estadística”.

El propio grupo de trabajo de inferencia estadística de la *American Psychological Association* (Wilkinson and the Task Force on Statistical Inference, 1999) propone en su informe final acompañar las pruebas de significación estadística con otros métodos estadísticos como el tamaño del efecto y sus intervalos de confianza o la aplicación de la estadística bayesiana. Es esencial, en opinión del grupo de trabajo, informar del tamaño del efecto y su intervalo de confianza e interpretarlo dentro del contexto de los valores de tamaño del efecto obtenidos en investigaciones previas (p. 599). Dicho informe provocó cambios importantes en las recomendaciones de la quinta edición del Manual de Publicación de la *American Psychological Association* (2001), destacando la necesidad de acompañar *siempre* los resultados del trabajo con la estimación del tamaño del efecto o de la magnitud de la relación entre las variables. Como consecuencia, las políticas editoriales de las revistas también incluyen nueva normativa sobre los análisis estadísticos y se han creado líneas de investigación específicas para el desarrollo de software que contenga las técnicas que los grandes paquetes estadísticos aún no han incluido en los menús de su software como es el caso del SPSS. El siglo XXI comienza con un espíritu de reforma estadística que paulatinamente va cobrando fuerza y presencia en las publicaciones científicas.

La reforma estadística cambia el punto de mira desde “*cuán probable o improbable es el resultado muestral*” hacia dos cuestiones principalmente: “*cómo de grande es el tamaño del efecto detectado*” y “*si se puede replicar*”. Es decir, hay que ‘evaluar’ el valor del tamaño del efecto estimado y situarlo en el contexto más amplio de otras investigaciones de la misma temática comparado e integrando los resultados de forma explícita y directa. Hay que tener en cuenta que la cuantía del tamaño del efecto remite a su ‘no trivialidad’, pero no necesariamente; por ejemplo, un efecto pequeño en reducción de muertes puede ser más importante que otro más grande en reducción de

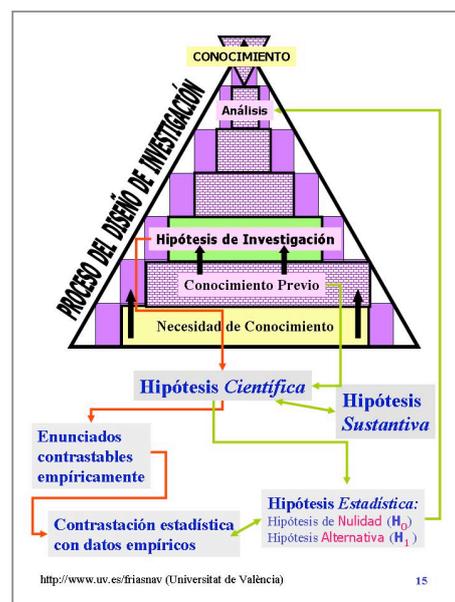
grasa corporal. Además, la replicabilidad del efecto supone evaluar cuán estables son los efectos en la literatura revisada y por lo tanto evaluar en qué medida son efectos directamente comparables. En definitiva, el nuevo comportamiento del investigador supone desarrollar el denominado “*pensamiento meta-analítico*” (Cumming y Finch, 2002).

Sin embargo, la conducta de los investigadores no llega a consolidar las prácticas de la nueva práctica estadística (Vacha-Haase, Nilsson, Reetz, Lance y Thompson, 2000). Seguramente los investigadores aún no han reflexionado sobre la cuestión de que “*seguramente Dios ama al 0.06 (nivel de significación estadística) tanto como al 0.05*” (Rosnow y Rosenthal, 1989, p.1277).

Modelo tradicional de decisión estadística: procedimiento de significación de la hipótesis nula (NHST)

En gran parte de las ocasiones, la meta principal de la investigación es hacer una inferencia sobre un efecto en una población a partir de los datos de una muestra que sólo nos interesa como indicador de lo que ocurre en la población. Se trata de un proceso de inferencia estadística. El científico no puede verificar una teoría pero sí refutarla. El modelo tradicional de decisión estadística implica definir dos hipótesis estadísticas que entran en competencia: la denominada *Hipótesis Nula* (H_0) o de nulidad de efectos que típicamente mantiene que no hay efecto del tratamiento (efecto cero) o ausencia de relación entre las variables (correlación cero) y la *Hipótesis Alternativa* (H_1) que postula que sí hay efecto (efecto diferente a cero) o sí existe relación entre las variables (correlación diferente de cero) (ver Tabla 1).

Cohen (1994) llamó a la hipótesis nula de efectos cero la ‘*nil hypothesis*’ frente a otras hipótesis nulas de efectos diferentes de



cero (*'non-nil hypothesis'*). Una *'nil hypothesis'* plantea que la correlación o la diferencia de medias es literalmente cero.

Tabla 1. Hipótesis Nula/Hipótesis Alternativa

HIPÓTESIS NULA	HIPÓTESIS ALTERNATIVA
Asume que no hay diferencias entre las poblaciones de los grupos. Es decir, el efecto en la población es cero. Por lo tanto: 1) Las pequeñas diferencias encontradas en las poblaciones serán debidas al azar 2) Las grandes diferencias ocurrirán en muy pocas ocasiones y serán también por azar	Hay diferencias entre las poblaciones de los grupos. El efecto en la población es diferente de cero. Por lo tanto: 1) Las pequeñas o grandes diferencias están provocadas por el efecto del tratamiento

Las hipótesis estadísticas son afirmaciones sobre parámetros de la población como por ejemplo una media ($H_0: \mu_1 = \mu_2$) o una proporción y no son hipótesis sobre una muestra. La meta del procedimiento de decisión estadística es por lo tanto evaluar si se puede o no rechazar la hipótesis nula (H_0) con un cierto grado de plausibilidad. En el caso de rechazar la hipótesis nula entonces se podrá aceptar la hipótesis alternativa y aquí comienzan ya una de las críticas más fuertes al procedimiento de significación estadística: aportar conclusiones sobre la hipótesis alternativa cuando es la hipótesis nula la que se postula en la decisión estadística.

Errores estadísticos

Dentro del modelo de decisión estadística se trabaja con un cierto grado de incertidumbre estadística elegido por el investigador a priori mediante un criterio de decisión probabilística conocido como *alfa* (α) o error de Tipo I (*'probabilidad de rechazar la hipótesis nula siendo realmente verdadera'*, es decir, cuando la diferencia entre las poblaciones está realmente provocada por el azar) (ver Tabla 2.). Es el nivel de significación estadística.

Tabla 2. Errores estadísticos y decisiones correctas

Decisión estadística	Realidad del fenómeno	
	H ₀ cierta $p(H_0) = 1$	H ₁ cierta $p(H_0) = 0$
Mantener la Hipótesis Nula	Nivel de confianza (1 - <i>alfa</i>)	Error de Tipo II (<i>beta</i> , β)
Rechazar la Hipótesis Nula	Error de Tipo I (<i>alfa</i> , Nivel de significación)	Potencia estadística (1 - <i>beta</i>)

Tradicionalmente el nivel de *alfa* máximo se ha fijado en 0.05 o en el 5%, es decir, cinco veces de cada cien se rechazará la hipótesis nula siendo realmente cierta. El resultado de $1 - \textit{alfa}$ es la ‘*probabilidad de mantener la hipótesis nula cuando realmente es verdadera*’ (por ejemplo es el nivel de confianza de 0.95 cuando *alfa* es 0.05).

El riesgo de no rechazar la hipótesis nula cuando realmente es falsa se conoce como *beta* (β) o error de Tipo II (‘*probabilidad de mantener la hipótesis nula cuando realmente es falsa*’, es decir existe una diferencia real entre las poblaciones) y también se puede estimar evaluando la denominada ‘potencia estadística’ de la prueba estadística o la ‘*probabilidad de rechazar la hipótesis nula cuando realmente es falsa*’. La potencia estadística es $1 - \textit{beta}$. En general, desde un punto de vista de planificación de la investigación, el valor máximo de *beta* se sitúa en 0.20 (por lo tanto potencia estadística de 0.80). Sin embargo, a diferencia del valor de *alfa* (fijado a priori por el investigador) el investigador no tiene control directo del valor exacto de *beta* hasta que se ejecuta la prueba estadística y sólo puede planificarlo en función de una serie de parámetros como tamaño del efecto, número de observaciones y *alfa*. De ahí que cualquier cambio que se produzca en los datos respecto a tamaño del efecto y número de observaciones afectará al valor de potencia estadística a posteriori, es decir el obtenido en la investigación. Dependiendo del contexto de investigación, el tamaño del efecto puede ser la diferencia entre medias de dos grupos o la correlación entre dos variables.

Como se observa, en el proceso de decisión estadística se admite un mayor grado de error de Tipo II que de Tipo I, en concreto cuatro veces más de error *beta* que de *alfa*. Es preferible no concluir que existe un efecto, aunque sí exista, que hacer afirmaciones de efectos que realmente no se han producido.

La decisión de rechazar o no rechazar la hipótesis nula se realiza comparando la probabilidad de los datos obtenidos en la investigación bajo el modelo de la hipótesis nula o de efectos cero (valor p de probabilidad) con el nivel de *alfa* elegido por el investigador. **Si la probabilidad de los datos obtenidos en la investigación, asumiendo que la hipótesis nula es cierta, es menor o igual al valor de incertidumbre asumido a priori en el diseño de investigación (nivel de *alfa*) entonces se rechaza la hipótesis nula con ese margen de error o equivocación en la decisión estadística.**

Por lo tanto, el valor p es la probabilidad del resultado obtenido (o un resultado más extremo) asumiendo que la hipótesis nula es cierta. Así:

- Si $p = 1$ entonces 'es completamente cierto' que el dato obtenido pertenece a una distribución que plantea la igualdad de las poblaciones (hipótesis nula).
- Si $p = 0$ entonces 'es absolutamente imposible' que el dato obtenido pertenezca a la distribución que plantea la igualdad de las poblaciones (hipótesis nula) y por lo tanto pertenecerá a la hipótesis alternativa que mantiene la diferencia entre las poblaciones.

Evidentemente no hay dos grupos que sean exactamente iguales antes del tratamiento. Incluso aunque se utilice la asignación aleatoria existirán diferencias individuales o cuestiones de muestreo que provocarán cierta variabilidad. La técnica estadística determina la probabilidad de que las diferencias observadas entre los grupos podrían ser debidas al mismo proceso de asignación aleatoria de las unidades experimentales que forman los grupos. Si la probabilidad (valor p) de encontrar dichas diferencias es baja entonces se asume que las diferencias entre los grupos están causadas por el efecto del tratamiento y por lo tanto se rechaza la hipótesis nula. Si la probabilidad de encontrar diferencias debido al proceso de asignación aleatoria es alta entonces la explicación del azar es la mejor elección y por ello se mantiene la hipótesis nula o de nulidad de efectos en la población. **Decidir qué es una probabilidad alta o una probabilidad baja depende del valor que se seleccione a priori como *alfa*. Se considera que la probabilidad es baja cuando p es menor o igual a *alfa* y se considera que la probabilidad es *alta* cuando el valor de p es mayor al de *alfa*.**

En definitiva, la decisión estadística (*Mantener* H_0 o *Rechazar* H_0) supone valorar si el

valor del efecto observado en la muestra es improbable bajo el modelo de la hipótesis nula en cuyo caso se rechaza la hipótesis nula. El valor de improbable (o probable) lo valoramos al comparar el valor p de probabilidad con el valor de α . De este modo:

-Si $p \leq \alpha$ entonces entra en la categoría de improbable bajo el modelo de la hipótesis nula, asumiendo cierto error de Tipo I. Rechazamos la hipótesis nula y por lo tanto existe un efecto diferente de cero. Si el valor de α es igual a 0.05 entonces diremos que el efecto es estadísticamente significativo al 0.05 o al 5%.

-Si $p > \alpha$ entonces no hay suficiente evidencia para rechazar la hipótesis nula y por lo tanto mantenemos que el resultado pertenece a la distribución de dicha hipótesis hasta que se demuestre lo contrario. Las diferencias encontradas es muy probable que se deban al azar. Se mantienen la hipótesis nula, es decir, el efecto encontrado no es estadísticamente significativo.

La denominada ‘significación estadística’ estima la probabilidad, valor p , de los resultados muestrales dado el tamaño de la muestra y asumiendo que dicha muestra procede de una población descrita por la distribución de la hipótesis nula (Cohen, 1994; Thompson, 1996). Por ejemplo, tenemos dos muestras con dos medias diferentes en una determinada variable. El investigador puede conocer si esa diferencia es sólo azarosa o ha sido producida por algún proceso sistemático. Si las dos muestras fueron configuradas a partir de una misma población donde el parámetro de interés es distribuido normalmente, las leyes de la probabilidad permiten cuantificar la probabilidad condicional de la diferencia en los valores de los parámetros muestrales. Si la diferencia es muy improbable el investigador tiene dos opciones: concluir que ha ocurrido una diferencia improbable o concluir que la diferencia tiene otras causas además de la variación aleatoria. El rechazo de la hipótesis nula opta por la segunda opción, sujeta también a error dado que podría tratarse únicamente de variación muestral aleatoria (Error de Tipo I).

Por lo tanto, y siguiendo el modelo tradicional de comprobación de hipótesis, cuando se ejecuta la prueba de significación estadística, la *hipótesis científica* gana o pierde credibilidad a partir de los resultados de **un contraste estadístico que somete a prueba el rechazo o no rechazo de la hipótesis nula, asumiendo desde el primer momento**

que es una hipótesis cierta. El rechazo de la hipótesis nula supondrá hacer inferencias sobre qué hipótesis alternativa es la que mejor explica los hallazgos. Si en el diseño de investigación se han controlado adecuadamente las amenazas a la validez de los hallazgos habrá que suponer que los efectos se deben a la hipótesis alternativa formulada en el estudio (siempre con un margen de error de Tipo I). Nunca se somete a contraste estadístico la hipótesis alternativa y es este punto el que inicia una línea de críticas especialmente duras al proceso clásico de inferencia estadística junto con la escasa credibilidad que la hipótesis nula tiene para el investigador (tamaño del efecto de cero).

Hasta aquí se ha presentado el procedimiento estándar que se utiliza para evaluar empíricamente las hipótesis. Sin embargo, la interpretación de los resultados aportados por la prueba de significación estadística no es unánime por parte de los investigadores. Realizado el proceso de decisión estadística falta que el investigador interprete los hallazgos y aquí sus creencias y atribuciones sobre el significado y alcance de los resultados no siempre son correctas. El factor humano se convierte en tema de debate sobre los usos y abusos por parte de los investigadores de las pruebas de significación estadística (Monterde, Pascual y Frías, 2006; Oakes, 1986).

Pongamos un ejemplo. Siguiendo el estudio clásico de Oakes (1996), se presentó a una muestra de investigadores un listado de siete afirmaciones sobre la interpretación de los resultados del contraste estadístico donde se obtuvo que el valor de $p = 0.01$ ($p < \alpha$). Cada uno de ellos debía valorar el listado de afirmaciones siguiente como verdadera o falsa:

- a) La hipótesis nula es absolutamente rechazada
- b) Se ha determinado la probabilidad de la hipótesis nula
- c) La hipótesis alternativa es absolutamente rechazada
- d) Se ha deducido la probabilidad de la hipótesis alternativa
- e) Se conoce la probabilidad de que sea errónea la decisión
- f) Una replicación posterior tendría 0.99 de probabilidad de ser significativa
- g) La probabilidad de obtener ese dato o más extremo si la hipótesis nula es cierta

De las seis afirmaciones anteriores sólo la última es cierta. La confusión entre la

probabilidad de los datos asumiendo una hipótesis como cierta ($p(D|H)$) y la probabilidad de la hipótesis dados ciertos datos ($p(H|D)$) conduce a las interpretaciones erróneas reflejadas en las opciones *a*, *b*, *c*, *d* y *e*. En la opción *f* se confunde el valor de p con la potencia estadística.

El valor p de probabilidad informa de la probabilidad de los datos obtenidos dado que la hipótesis nula es cierta ($p(D|H_0)$). Sin embargo, muchos investigadores siguen creyendo erróneamente (opción *b* de las afirmaciones anteriores) que el valor p de probabilidad es la probabilidad de que la hipótesis nula sea cierta dados los datos de la investigación ($p(H_0|D)$). Una prueba de significación estadística ofrece la probabilidad posterior de los datos **condicional a la verdad de la hipótesis sometida a contraste (la hipótesis nula)**. Por ello, las pruebas de significación no son apropiadas cuando las muestras no han sido aleatoriamente configuradas de la población. Las pruebas de significación estadística no ofrecen información de la probabilidad condicional de la hipótesis dados los datos obtenidos en la investigación.

Erróneamente muchos investigadores creen que el valor p de probabilidad es el error de Tipo I o *alfa* ‘observado’ en los datos. Sin embargo, no es cierto. El valor p de probabilidad es una medida de evidencia contra la hipótesis nula y no es una medida de apoyo a la hipótesis alternativa. El valor p depende de los datos del experimento, no es un valor fijo mientras que el valor de *alfa* no depende de los datos y sí es un valor fijo que determina a priori el investigador.

En definitiva, los problemas de interpretación del alcance del proceso de decisión estadística corresponden al factor humano y no a la prueba estadística como procedimiento metodológico. Es tarea de los docentes y de los manuales de metodología explicar con precisión los conceptos implicados en la prueba de significación estadística.

Sin embargo, el proceso de decisión estadística sí está sujeto a cuestiones metodológicas que afectan a la validez de los resultados. Es el caso del tamaño de la muestra o de la improbabilidad lógica de la hipótesis nula y su falta de plausibilidad. Cuestiones que abordaremos en otro momento.

Errores humanos implicados en el proceso de significación estadística

Es cierto, resulta extremadamente difícil para los alumnos comprender el proceso de decisión estadística y lo que realmente es la significación estadística y el valor p de probabilidad. Y no sólo para los alumnos sino también para los profesores y profesionales de la Psicología. Los procesos cognitivos implicados en la dificultad de comprensión de los conceptos que forma parte del proceso de decisión estadística han sido analizados durante años. Muchas estrategias de aprendizaje se han desarrollado para mejorar la comprensión pero los resultados no son alentadores. Una creencia errónea muy común es confundir el valor p de probabilidad con el valor de *alfa*.

Existe una gran confusión entre el valor p de probabilidad y el valor de *alfa* (error de Tipo I) del diseño. El valor p de probabilidad se calcula con los datos de la investigación que se están analizando en ese momento y refleja la probabilidad de los datos suponiendo que la hipótesis nula es cierta. Es un dato que sólo se conoce a posteriori, es decir, cuando se ejecuta la prueba estadística y por lo tanto está vinculado al valor obtenido con dicha prueba y los datos e indica la probabilidad de encontrar un resultado de la magnitud hallada en la investigación (o mayor), **condicional siempre a que la hipótesis nula sea cierta**. Para calcular el valor p de probabilidad se necesita conocer la distribución de la prueba estadística bajo el modelo de la hipótesis nula (representada típicamente en las tablas de los manuales de metodología) y los grados de libertad del diseño de investigación aplicado. El valor de *alfa* o error de Tipo I se fija a priori, antes de ejecutar la investigación, y es una decisión que toma el investigador en función de la planificación de su diseño de investigación. Ese valor de *alfa* es la referencia que se utiliza como valor de comparación con el valor p de probabilidad obtenido en el estudio para poder tomar la decisión estadística. El valor p de probabilidad ni es la significación estadística ni tampoco señala la magnitud mayor o menor del efecto.

El valor p de probabilidad es una medida de evidencia contra la hipótesis nula desde el modelo de Fisher, de tal modo que cuanto más pequeño sea su valor mayor será la evidencia. Desde el modelo de Neyman y Pearson se acompaña la hipótesis nula con la hipótesis alternativa y ahora hay que tomar la decisión de si se mantiene H_0 o en cambio

se rechaza a favor de H_1 . Aparecen entonces los conceptos de error de Tipo I y error de Tipo II. A partir de los dos modelos se crea un híbrido que se presenta en los manuales como el *procedimiento de significación estadística* donde el investigador realiza las siguientes consideraciones:

- 1º. Formula la hipótesis nula (H_0) y la hipótesis alternativa (H_1). El procedimiento comienza con la **asunción de que la hipótesis nula es cierta**. Sin embargo, realmente la meta es determinar si hay suficiente evidencia para inferir que la hipótesis alternativa es cierta
- 2º. Establece el nivel de significación (*alfa*) o el error de Tipo I a priori. Además supuestamente se estima la potencia de la prueba estadística (*1-beta*) teniendo en cuenta el tamaño del efecto esperado y el tamaño de la muestra
- 3º. Se computa la prueba estadística seleccionada por el investigador y se obtiene el valor p de probabilidad vinculado al valor del estadístico calculado
- 4º Finalmente se toma una decisión estadística que conlleva cometer uno de los dos errores estadísticos: el error de Tipo I (rechazar una hipótesis nula verdadera) o el error de Tipo II (no rechazar una hipótesis nula falsa). Se establece la 'significación estadística' utilizando el criterio de que si $p \leq \alpha$ ese resultado es estadísticamente significativo con el margen de error de *alfa* y si $p > \alpha$ entonces el resultado no es estadísticamente significativo

“TESTIGO DE CARGO”

Vamos a utilizar la analogía entre la decisión del jurado en un juicio y el proceso de decisión estadística para ofrecer una comparación con un modelo de decisión que todos



conocemos aunque sea por las películas o los documentales. Se va a utilizar como ejemplo de comparación la clásica película de Billy Wilder “Testigo de Cargo” (1957, *Witness for the Prosecution*) donde un afable personaje, Leonard Vole, es acusado del asesinato de una rica dama, la señora French, con quien mantenía una relación de carácter amistoso. El posible móvil del crimen es la herencia de una importante cantidad de dinero de la difunta. A pesar de que las pruebas en su contra son demoledoras, el prestigioso abogado criminalista de Londres Sir Wilfrid Roberts acepta su defensa al creer en su inocencia. Resulta imprescindible que el lector haga un visionado completo de la película especialmente del momento donde se conocen

abogado y acusado (primera parte de la película) y del final de la película para abordar como conocimiento la analogía del proceso de decisión estadística.

En un juicio con jurado el investigador actúa como el fiscal o como acusación particular y su objetivo es demostrar la falsedad de la hipótesis nula y como consecuencia la aceptación de la hipótesis alternativa (*culpabilidad del acusado*). Los miembros del jurado y el juez son los encargados de evaluar la calidad de la evidencia aportada por el investigador. En el proceso de investigación científica, la calidad de la evidencia aportada por un estudio es evaluada por otros investigadores, los denominados revisores de las revistas, los editores o los mismos lectores. El abogado defensor actúa como revisor de la evidencia o pruebas aportadas por el fiscal tratando de mostrar los sesgos y deficiencias de dichas pruebas (validez), defendiendo la hipótesis nula (*inocencia del acusado*). El fiscal actúa como defensor de la hipótesis alternativa y su tarea es aportar evidencias que apoyen la culpabilidad del acusado (presencia del efecto de tratamiento).

Antes de iniciar un juicio es necesario que el juez valore que existe suficiente evidencia preliminar que incrimina al acusado (hipótesis alternativa) y por ello es necesario iniciar un juicio. La hipótesis alternativa es la razón por la que se arresta al acusado. Por ejemplo, para permitir una orden de arresto, el juez debe encontrar que hay una causa probable de que la persona ha cometido un crimen. Hay un móvil. Del mismo modo, para que un policía arreste a un individuo sin una orden judicial deberá existir una causa razonable para creer que un crimen es inminente o que se acaba de cometer. En el proceso de investigación científica se parte de una necesidad de conocimiento unida a una hipótesis sustantiva (generalmente vinculada a la hipótesis alternativa) que inicia el proceso de diseño estadístico.

Todo proceso judicial comienza con la asunción de que el acusado es inocente hasta que se demuestre lo contrario. Desde la fiscalía o desde la acusación particular se aportarán todas las pruebas que se encuentren durante el desarrollo del juicio y tengan calidad para demostrar que el acusado es culpable más allá de toda duda razonable (siempre podrán existir explicaciones o hipótesis alternativas pero serán muy improbables). Al final del juicio, el jurado y el juez tomarán una decisión: el acusado es *No Culpable* o el

acusado es *Culpable* (ver tabla 3). Téngase en cuenta que en un juicio nunca se concluye que el acusado es inocente, el veredicto es *No Culpable*. Es decir, no se han encontrado pruebas de suficiente calidad para poder declararlo culpable. En un proceso de decisión estadística **nunca se concluye que la hipótesis nula es cierta**, sólo se mantiene dado que hasta el momento la evidencia encontrada no permite rechazarla.

Tabla 3. Sentencia y condición del acusado

Sentencia del juez	Condición real del acusado	
	Inocente	Culpable
No Culpable	Decisión correcta	Decisión incorrecta
Culpable	Decisión incorrecta	Decisión correcta

Por lo tanto, si las pruebas aportadas en el juicio no son suficientes para probar la culpabilidad más allá de una duda razonable entonces el acusado será declarado *No Culpable*. El acusado es declarado *Culpable* cuando hay suficiente evidencia para probar su culpabilidad más allá de una duda razonable. En este caso, se podrá apelar el veredicto.

El proceso de investigación científica se inicia porque previamente existe una justificación teórica que ha provocado una hipótesis sustantiva de investigación que desea ser contrastadas con un rigor metodológico (muestreo adecuado, planificación de la potencia estadística, elección del diseño). Hipótesis teórica que se representa bajo la hipótesis alternativa. Inicialmente, en el proceso de decisión estadística la hipótesis nula se asume como cierta (presunción de inocencia del acusado) hasta que se demuestre su falsedad (su rechazo) más allá de una duda razonable (margen de error de Tipo I).

En el proceso judicial y en el de decisión estadística no es posible probar la verdad de forma absoluta o con total certeza. En el juicio el criterio de decisión es “una duda razonable”. Si hay una duda razonable sobre la culpabilidad del acusado entonces será declarado no culpable. Pero si hay evidencia que señalan su culpabilidad más allá de una duda razonable entonces el acusado será declarado culpable. En el proceso de decisión estadística la hipótesis nula se mantiene como principio mientras no exista una duda razonable de su falsedad. Es decir, se asume como principio que la causa del

efecto detectado en los datos es la variabilidad aleatoria y no la causa que está siendo investigada por el investigador. El criterio para cambiar de decisión se fija con el nivel de *alfa* o probabilidad de Error de Tipo I. Por lo tanto, el ‘más allá de una duda razonable’ viene marcado por la probabilidad de aceptar como verdadera la hipótesis alternativa cuando no lo sea (hipótesis nula realmente cierta) con un margen de error del 5% si se trabaja con alfa de 0.05. Es decir, obtener un resultado estadísticamente significativo indica que se ha encontrado suficiente evidencia para probar que la hipótesis nula no es verdadera más allá de una duda razonable (5% o 1%), proporcionando apoyo a la hipótesis alternativa. Un resultado estadísticamente no significativo puede estar provocado por la falta de potencia estadística. Quizás el fiscal no encontró las mejores pruebas o las pruebas más contundentes para declarar culpable al acusado.

Por lo tanto, el acusado puede ser declarado No Culpable o Culpable; la hipótesis nula puede ser que se Mantenga o se Rechace. La hipótesis alternativa refleja que el acusado es culpable y supone rechazar el modelo de la hipótesis nula y como consecuencia aceptar la hipótesis alternativa. Y esto es una decisión que sólo puede tomarse en función de la evidencia encontrada gracias al diseño de investigación y la validez de los hallazgos. El jurado sólo podrá declarar culpable al acusado cuando haya suficiente evidencia o pruebas de calidad que apoyen su culpabilidad.

Lamentablemente ni el sistema judicial ni la decisión estadística son perfectos. Un jurado puede declarar Culpable a un inocente (ver tabla 3) o una decisión estadística puede implicar rechazar la hipótesis nula cuando realmente es cierta. Se trata del error de Tipo I.

También puede ocurrir que el jurado no rechace la presunción de inocencia porque no se aportaron suficientes pruebas de culpabilidad (no porque de verdad sea inocente). El veredicto de *No Culpable*, no equivale a inocente porque la inocencia no se prueba, sólo se llega a la conclusión de que con las pruebas aportadas el acusado no puede ser declarado culpable. Quizás sea culpable, pero con la información disponible el veredicto es *No culpable*. Conviene tener en cuenta este aspecto porque la hipótesis nula tampoco se puede probar, sólo se concluye que no se puede rechazar (o sí se puede) pero nunca

se dirá que se acepta la hipótesis nula. No se demuestra la verdad de la hipótesis nula.

Otro problema de las decisiones judiciales es declarar no culpable a una persona realmente culpable. Una decisión estadística puede implicar mantener la hipótesis nula cuando realmente es falsa. Se trata del error de Tipo II o beta. En general, el máximo de error de Tipo II está fijado en 0.20, es decir, es cuatro veces mayor que el error de Tipo I permitido. Podríamos decir que es mucho más grave declarar culpable a un inocente (error de Tipo I) que inocente a un culpable (error de Tipo II). En la Tabla 4 se resumen los elementos que permiten comparar el proceso de juicio y el proceso de decisión estadística. El alumno debe completar la tabla de analogía.

Tabla 4. Analogía entre el proceso de juicio y el de decisión estadística

	PROCESO DE JUICIO	PROCESO DE DECISIÓN ESTADÍSTICA
Inicio	Asunción de inocencia del acusado	Asunción de que la hipótesis nula es cierta
El objetivo del juicio		
El criterio para determinar la culpabilidad		
Previo al proceso		
Durante el proceso		
Veredicto/Decisión		
Decisión correcta		
Decisión incorrecta		
Calidad de la decisión	Calidad de las pruebas aportadas	

Recortar las siguientes instrucciones y pegarlas en el portafolio. Después completar la práctica. -----

Trabajo de portafolio:

- 1.- El alumno deberá realizar una reflexión sobre el proceso de decisión estadística: en qué consiste, sus pasos y su interpretación.
- 2.- Posteriormente completará la tabla de analogía y razonará sus decisiones (tabla 4).
- 3.- A continuación recogerá las escenas de la película ‘*Testigo de cargo*’ que considere válidas para ejemplificar cada uno de los elementos del proceso de decisión estadística. Puedes transcribir los diálogos o describir las escenas junto con su momento paralelo en el proceso de decisión estadística.
- 4.- Se completará un apartado final donde se indicarán las horas dedicadas al trabajo propuesto incluyendo el tiempo de visionado de la película. Y se cumplimentará el siguiente cuestionario: RECORTAR Y PEGAR EN EL PORTAFOLIO.

-----RECORTAR Y PEGAR-----

HORAS DEDICADAS AL TRABAJO DE PORTAFOLIO DE LA PELÍCULA “TESTIGO DE CARGO” (incluido el tiempo de los visionados de la película)										_____ HORAS
Dificultad de la tarea propuesta en el portafolio	Nada 1	2	3	4	5	6	7	8	9	Mucho 10
Satisfacción con el trabajo realizado	Nada 1	2	3	4	5	6	7	8	9	Mucho 10
Comprensión de los conceptos metodológicos	Nada 1	2	3	4	5	6	7	8	9	Mucho 10
Calidad de la película elegida como ejemplo de la analogía	Mala 1	2	3	4	5	6	7	8	9	Buena 10
Calidad de los apuntes teóricos que acompañan a la práctica	Escasa 1	2	3	4	5	6	7	8	9	Buena 10
¿Crees que vale la pena incluir esta práctica para comprender el proceso de decisión estadística? Tacha la respuesta correcta según tu opinión personal	SÍ	NO LO TENGO CLARO								NO

5. El alumno señalará qué conceptos le han quedado más claros con el ejemplo de la analogía. Es muy importante que el alumno señale qué conceptos conocía de forma errónea y definirlos erróneamente tal y como pensaba antes y a continuación cómo los entiende ahora. Es importante que anote el recurso de la analogía que le ha permitido avanzar en su comprensión.

7. Finalmente se realizará una valoración personal de la experiencia y se apuntarán nuevas tareas o ejemplos para mejorar la comprensión del proceso de decisión estadística.
