

# NTP 401: Fiabilidad humana: métodos de cuantificación, juicio de expertos



Fiabilité humaine: méthodes de quantification, jugement d'experts  
Human reliability: quantification methods, experts judgement

Vigencia	Actualizada por NTP	Observaciones	
Válida			
ANÁLISIS			
Criterios legales		Criterios técnicos	
Derogados:	Vigentes:	Desfasados:	Operativos: <b>SI</b>

## Redactora:

M. Isabel de Arquer  
Licenciada en Psicología

CENTRO NACIONAL DE CONDICIONES DE TRABAJO

El objetivo de esta NTP es exponer, brevemente, una forma alternativa de obtener información cuantitativa de utilidad para los análisis de fiabilidad humana. Complementa a las N.T. P. 360-94 «Fiabilidad humana: conceptos básicos» y 377-95 «Fiabilidad humana: métodos».

## Introducción

Los análisis de fiabilidad humana desde una perspectiva cuantitativa precisan datos acerca de la consistencia de respuesta personal (probabilidad de que ante los mismos conjuntos de estímulos, se desencadene el mismo o similar tipo de respuesta). Cifrar la probabilidad de error humano no es nada sencillo y resulta especialmente difícil en algunas tareas y circunstancias.

La obtención de información y la elaboración de bases de datos útiles para los análisis cuantitativos de fiabilidad humana presentan problemas ocasionados, entre otras razones, porque:

- normalmente, sólo se suele informar de los errores humanos que se traducen en consecuencias no deseables; sería conveniente asegurar la notificación y el registro de todos los errores, incluso de los que no se siguen de resultados negativos, con el objetivo final de aumentar la fiabilidad del sistema;
- la posibilidad de generalizar los datos plantea limitaciones a causa de las diferencias entre unas situaciones y otras, las distintas clasificaciones de tareas y tipos de acciones, así como la variedad en cuanto al número y tipo de factores considerados con influencia sobre la conducta; todo ello puede dificultar la aplicabilidad de los datos de un sistema a otros;
- para obtener datos fiables sobre las acciones erróneas de baja frecuencia pueden ser necesarios períodos de observación muy grandes;
- en ocasiones puede ser muy difícil definir el límite exacto entre la conducta humana adecuada y la errónea.

Las estimaciones y los juicios dados por expertos se presentan como una opción que hace posible obtener algunos de los valores cuantitativos sobre fiabilidad humana que se requieren para el análisis probabilístico de riesgos de un sistema.

## Juicios de expertos

Una de las técnicas aplicables en el análisis de fiabilidad humana es la estimación de probabilidad de error humano por medio de juicios de expertos. Se recurre a ella cuando, para conocer la probabilidad de error humano, hay que considerar la influencia de muchos factores interdependientes que afectan a la respuesta de la persona; también se emplea cuando en las bases de datos no se encuentran puntos de referencia aplicables.

Disponer de una base de datos sobre errores humanos constituye un buen punto de partida, pero puede no ser suficiente y presentar vacíos informativos. Entonces, las opiniones o juicios de los expertos son una alternativa que combina las aproximaciones analíticas y los métodos de estimación subjetiva de probabilidades desarrollados conforme a las reglas de la teoría de la decisión.

Los expertos se pueden pronunciar sobre los índices de error que estiman para actividades que no estén contempladas en la base de datos que se tiene. También pueden ofrecer sus reflexiones acerca de la influencia de determinados factores sobre el comportamiento de las personas, para así reflejar adecuadamente las diferencias entre los valores que se tengan de una base de datos y la situación y sistema reales que se pretende estudiar (ver cuadro 1).

## Cuadro 1: Razones para utilizar el juicio de expertos

LA CONSULTA A EXPERTOS PERMITIRÍA OBTENER LA PROBABILIDAD ESTIMADA DE ERROR HUMANO EN DIVERSAS ACTIVIDADES O TAREAS PARA LAS QUE:

- NO SE TIENEN DATOS DE REFERENCIA A PRIORI.
- HAY MUCHOS FACTORES CUYA INFLUENCIA SOBRE LA CONDUCTA DEBE SER PONDERADA.
- SE NECESITA AJUSTAR A UNA REALIDAD, LOS VALORES DE REFERENCIA DE UNA BASE DE DATOS (CUANDO SE DISPONE DE ELLOS) DE UN SISTEMA DISTINTO.

En todo caso, mediante el juicio de expertos se pretende tener estimaciones razonablemente buenas, las «mejores conjeturas», a faltade cifras más exactas. Sin embargo, estas estimaciones pueden y deben ser confirmadas o modificadas a lo largo del tiempo, según se vaya recopilando información durante el funcionamiento del sistema.

En un estudio comparativo de diversas técnicas y métodos al uso en el campo de la fiabilidad humana se califica el método de juicios de expertos como un método aceptable en la comunidad científica, de una elevada madurez y que proporciona estimaciones de precisión moderada. Su validez y su utilidad se califican entre moderadas y altas frente al calificativo de moderadas para el método THERP (Technique for human error rate prediction).

## Juicio de expertos: proceso

Inicialmente, hay que tener una idea clara acerca de qué datos se necesitan. Puede ocurrir que el conjunto de datos de que se dispone sea incompleto o insuficiente de algún modo. Puede que se ignoren las condiciones en las que se han obtenido tales datos o puede desconocerse si la situación a la que queremos aplicarlos es suficientemente similar a aquella que los originó.

Como requisito previo se necesita disponer de una definición y descripción claras de las tareas y actividades sobre las cuales los expertos deberán emitir un juicio en términos de probabilidades de éxito o fracaso. Cuanto más específica sea la definición de una tarea o de una actividad, menos margen quedará para posibles interpretaciones erróneas. A veces es recomendable consultar a varios especialistas para asegurarse de que la descripción de la tarea o actividad es clara y el nivel de detalle es suficiente, en concreto, en lo que se refiere a los factores que puedan afectar al nivel de ejecución de la persona que desempeña el trabajo.

Para seguir el proceso se debe seleccionar a los expertos. Pueden encontrarse personas «expertas» entre quienes desempeñan el trabajo, formadores, especialistas en seguridad, mandos, dirección y personas del equipo de proyectos y diseño. Para que una persona pueda ser etiquetada como «experta» debe poseer un conocimiento profundo de la tarea o actividad que será objeto de análisis y valoración y tiene que estar familiarizada con el sistema en el que ésta se desarrolla. Asimismo, los expertos a los que se recurre tienen que ser capaces de traducir su valoración en términos de probabilidades.

El número de expertos necesario para conseguir el juicio que se busca no es fijo, oscila desde tres o cuatro expertos hasta seis. Sin embargo, Villemeur menciona que en algunas ocasiones pueden utilizarse más de ocho expertos, según la precisión que se desee.

Se debe tener previsto el modo en que se recogerán los juicios de los expertos y tener preparados, si es el caso, cuestionarios, escalas de evaluación, etc. La precisión en éstos es esencial para que los resultados no queden sesgados por fallos de comprensión. También se deben preparar las instrucciones que se van a dar a los jueces o expertos, de modo que comprendan con claridad cuál es el objetivo de su ejercicio de evaluación (ver cuadro 2).

## Cuadro 2: Obtención de juicios de expertos: etapas de procedimiento

- ELABORAR UNA DESCRIPCIÓN DE LA/S TAREA/S.
- PREPARAR LAS INSTRUCCIONES PARA LOS JUECES/EXPERTOS.
- REALIZAR UNA PLANTILLA PARA RECOGER LAS RESPUESTAS DE LOS JUECES.
- SELECCIONAR A LOS EXPERTOS QUE SE HAN DE CONSULTAR.
- RECOGER LOS JUICIOS DE LOS EXPERTOS.
- CALCULAR LA CONSISTENCIA INTERJUECES.
- AGLUTINAR LAS ESTIMACIONES.
- CALCULAR LOS LÍMITES DE CONFIANZA.

## Métodos para la obtención de juicios de expertos

Los juicios de expertos se pueden obtener por métodos grupales o por métodos de experto único. Se pueden seguir, entre otros, el método de agregados individuales, el método Delphi, la técnica grupal nominal y el método de consenso grupal.

## Método de agregados individuales

Se pide individualmente a cada experto que dé una estimación directa de la probabilidad de éxito o de fracaso en cada una de las tareas descritas. Después se tratan estadísticamente los datos recogidos. Lo habitual es calcular la media aritmética del conjunto de estimaciones individualmente obtenidas, para cada tarea. Esto se hace presuponiendo que el conjunto de los datos posibles tiene una distribución simétrica y, por tanto, la media aritmética es un buen índice de tendencia central.

Éste es un método económico porque, al igual que el método Delphi, no exige que se reúna a los expertos en un lugar determinado. Puede parecer un método limitado porque los expertos no pueden intercambiar sus opiniones, puntos de vista y experiencia, ya que se les requiere individualmente; no obstante, esta limitación puede ser precisamente lo que se esté buscando para evitar los sesgos de los datos ocasionados por conflictos interpersonales, presiones entre los expertos, etc.

## Método Delphi

Este método fue creado en 1948 para obtener la opinión de expertos de una manera sistemática. En un primer momento, cada experto responde de manera individual y anónima a un cuestionario. Después se analizan las respuestas del conjunto de expertos, se remite a cada uno la respuesta mediana obtenida, así como el intervalo intercuartil para cada cuestión y se les pide que reconsideren su juicio anterior, teniendo en cuenta estos datos. En cada una de las tres o cuatro «vueltas» siguientes, se informa a los expertos de cuál es la mediana del grupo y se les propone revisar su juicio anterior. Todo juicio individual que quede fuera del intervalo intercuartil en que se mueve el grupo de expertos tiene que estar debidamente justificado o argumentado.

Una débil dispersión de los juicios obtenidos en la última vuelta indicaría que se ha alcanzado un consenso. La mediana de las respuestas obtenidas en esta última vuelta es el valor que se estaba buscando. Se emplea este estadístico de tendencia central (mediana), porque se presupone que las puntuaciones posibles de los expertos se distribuyen de forma asimétrica.

Con el método Delphi, los expertos comparten en cierto modo sus opiniones, sin que existan discusiones ni confrontaciones directas entre ellos.

## Técnica grupal nominal

El primer paso es reunir a los expertos (entre ocho y diez personas) y pedirles que registren, individualmente y sin intercambiar opiniones, sus propias puntuaciones y considerandos respecto a las probabilidades de error para cada una de las tareas/actividades que se les detallan. Después, cada experto expone a los demás las puntuaciones y principales considerandos registrados y al acabar esta ronda, todos los expertos disponen de la relación común del grupo. El paso siguiente consiste en un debate altamente estructurado de cada uno de los apartados de ésta. Finalmente, cada experto, de forma individual y por escrito, puntúa y argumenta las probabilidades de error para cada tarea/actividad considerada.

En general, se procede como con el método Delphi, con la salvedad de que en esta técnica se permite algún debate entre los expertos, para que aclaren y compartan la información que cada uno está considerando. No obstante, las discusiones son limitadas y así se limita también la presión de un/os expertos sobre los juicios de otros. Aunque hay intercambio de pareceres, los juicios se emiten de forma individual y la estimación final suele ser la media aritmética del conjunto de las estimaciones dadas por los expertos.

Para emplear esta técnica conviene que el grupo de expertos sea pequeño, se fomente la libre expresión y se eviten las críticas (discusiones tensas), el sistema de votación y el regateo. El éxito de la técnica depende, por una parte, de la habilidad y la experiencia de quien hace de moderador del grupo y por otra, de la buena voluntad del grupo de expertos para trabajar juntos en un marco altamente estructurado.

## Método del consenso grupal

Se necesita reunir a los expertos en un lugar determinado. Entonces se indica al grupo que su tarea consiste en lograr una estimación de la probabilidad de éxito o de fracaso para cada tarea, que sea satisfactoria para todos los expertos. Con estas instrucciones se maximizan los intercambios de información y opiniones dentro del grupo de expertos. Si el grupo no logra un consenso, puede intentarse un consenso artificial recogiendo las estimaciones individuales y sintetizándolas estadísticamente. Este método, como el anterior, también precisa que el grupo de expertos sea pequeño, se fomente la libre expresión y se eviten las discusiones tensas y los sistemas de votación.

## Resultados

Cuando no se obtiene de los expertos un valor consensuado en grupo, porque se ha optado, por ejemplo, por el método de expertos únicos, es necesario reunir las estimaciones individuales de los expertos.

Para conseguir un juicio significativo es necesario que haya una cierta homogeneidad entre los expertos en lo que se refiere a la comprensión de la tarea o actividad que han de valorar, las causas potenciales de errores y los factores que configuran la respuesta humana. Con el fin de obtener un resultado estadístico significativo, se calcula la consistencia interjueces. La excesiva heterogeneidad de los resultados no es aceptable, pues se corre el riesgo de dar una estimación que no sería representativa de nada; se debe realizar un análisis de varianza. La agrupación de los juicios de expertos se hace estadísticamente. La síntesis más común es una media geométrica, que efectúa una agrupación logarítmica de las opiniones que se han recogido en forma de porcentaje. Por último, se deben calcular los límites de confianza de los resultados obtenidos. Se presenta un sencillo ejercicio a modo de ejemplo.

Supongamos que el cuadro 3 recoge las estimaciones realizadas por cinco jueces (A, B, C, D y E), acerca de probabilidad de error

humano (HEP) para cada una de las cuatro actividades  $S_1$ ,  $S_2$ ,  $S_3$  y  $S_4$ .

**Cuadro 3: Estimaciones de probabilidad de error humano para varias actividades ( $S_1$ ,  $S_2$ ,  $S_3$ ,  $S_4$ ) según distintos jueces (A, B, C, D y E)**

Juez	$S_1$	$S_2$	$S_3$	$S_4$
A	,00200	,00100	,00020	,00010
B	,00500	,02000	,00050	,00002
C	,02000	,00200	,00020	,00002
D	,00200	,00500	,00010	,00005
E	,00100	,00050	,00050	,00010

Para calcular la consistencia interjueces se hace un análisis de varianza. Según las tablas de los puntos de significación de la distribución F de Snedecor (6), se tiene que para las columnas  $F_{0,001; 3, 12} = 10,80$  y para las filas  $F_{0,001; 4, 12} = 9,63$ ; comparan o estos valores con los obtenidos en el análisis de varianza del ejemplo ( $F_{columnas} = 15,4722$  y  $F_{filas} = 0,4119$ ) se puede afirmar que la influencia en el resultado debida al factor columnas es significativa a un nivel de confianza del 99,9%, mientras que la influencia debida al factor fila (juez) no es significativa (a un nivel de confianza del 99,9%). En definitiva, se puede considerar que, en el grupo de expertos que se ha consultado, hay una gran consistencia interjueces en cuanto a sus estimaciones de probabilidad de error humano para cada actividad ( $S_1$ ,  $S_2$ ,  $S_3$  y  $S_4$ ).

A continuación se calculan el error estándar y los límites de confianza de la estimación final de probabilidad de error humano para cada actividad. Para la actividad  $S_1$  la probabilidad de error humano es  $3 \cdot 10^{-3}$  y los límites son:  $5 \cdot 10^{-3}$  y  $1 \cdot 10^{-3}$ . Este resultado se puede considerar aceptable ya que, en general, el resultado será aceptable mientras el exponencial de los límites sólo varíe en un orden de magnitud. Para las otras actividades los resultados son:

- HEP ( $S_2$ ) =  $2 \cdot 10^{-3}$  (límites:  $4 \cdot 10^{-3}$  y  $1 \cdot 10^{-3}$ ).
- HEP ( $S_3$ ) =  $2 \cdot 10^{-4}$  (límites:  $3 \cdot 10^{-4}$  y  $1 \cdot 10^{-4}$ ).
- HEP ( $S_4$ ) =  $4 \cdot 10^{-5}$  (límites:  $6 \cdot 10^{-5}$  y  $3 \cdot 10^{-5}$ ).

## Ventajas y límites

Los resultados pueden estar afectados por algunos sesgos de origen diverso. Por ejemplo, la diferente facilidad para considerar los sucesos o errores más conocidos respecto de los errores menos conocidos, induce un sesgo en las estimaciones de los jueces o expertos.

Por otra parte, si los expertos trabajan en grupo, puede ocurrir que uno de ellos, por su especial forma de ser, arrastre a los demás hacia su propia opinión o a sobrestimar los criterios que él valora. También pueden surgir conflictos por motivos personales o de opinión que bloquean el trabajo del grupo, haciendo necesaria la intervención de una persona como mediadora para reanudar el diálogo.

El juicio probabilístico de expertos ha sido un recurso empleado con éxito para obtener estimaciones en diversos campos de actividad. Sus principales ventajas son: sus amplias posibilidades de aplicación a gran número de situaciones o casos y su sencillez de uso cuando se dispone de expertos bien informados y dispuestos a colaborar. Se puede utilizar en situaciones que comprendan tareas y actividades simples, actividades basadas en procedimientos, tareas de diagnóstico y de control de procesos, considerando el impacto de los factores de tipo socio-técnico sobre la conducta. Cuando para la obtención de los juicios de los expertos se emplea alguna técnica que permite el debate entre ellos se puede recoger información cualitativa útil para etapas posteriores, que posibilite la posterior reducción de la probabilidad de error en algunas tareas.

## Bibliografía

(1) HOLLNAGEL, E.

**Human reliability analysis. Context and control.**

Academic Press. London, 1993.

(2) INSTITUT DE SÛRETÉ DE FONCTIONNEMENT (ISdF),

**L'état de l'art dans le domaine de la fiabilité humaine.**

Octares Éditions. Toulouse, France, 1994.

(3) CENTER FOR CHEMICAL PROCESS SAFETY (CCPS)

**Guidelines for preventing human error in process safety.**

American Institute of Chemical Engineers (AIChE), New York, 1994.

(4) VILLEMEUR, A.  
**Reliability, availability, maintainability and safety assessment. Vol. 2.**  
John Wiley & Sons, Chichester, 1992.

(5) FINK, A., KOSECOFF, J., CHASSIN, M., BROOK, R.H.  
**Consensus methods: characteristics and guidelines for use.**  
Am. J. Public Health, 1984, **74**: 979-983.

(6) BOWKER, A.H., LIEBERMAN, G.J.  
**Méthodes statistiques de l'ingénieur.**  
Dunod, Paris, 1965.

(7) SEAVER, D.A., STILLWELL, W.G., SCHWARTZ, J.P.  
**Expert estimation of human error probabilities in nuclear power plants operations: a review of probability assessment and scaling.**  
NUREG/CR-2255, May 1982.

(8) HUMPHREYS, P.,  
**Human reliability assessors. Guide safety and reliability directorate.**  
United Kingdom Energy Authority. Wigshaw Lane, Culchetch Warrington, 1988.

(9) SALIOU, G.  
**Utilisation de jugements d'experts structurés.**  
EDF/DER/ESFHT 52/90-06A, 1990.