



TC/51/16

ORIGINAL: Inglés

FECHA: 11 de febrero de 2015

UNIÓN INTERNACIONAL PARA LA PROTECCIÓN DE LAS OBTENCIONES VEGETALES

Ginebra

COMITÉ TÉCNICO

Quincuagésima primera sesión Ginebra, 23 a 25 de marzo de 2015

REVISIÓN DEL DOCUMENTO TGP/8: PARTE I: DISEÑO DE LOS ENSAYOS DHE Y ANÁLISIS DE DATOS, NUEVA SECCIÓN: MINIMIZAR LA VARIACIÓN RESULTANTE DE LA EJECUCIÓN DE LOS ENSAYOS POR DISTINTOS OBSERVADORES

Documento preparado por la Oficina de la Unión

Descargo de responsabilidad: el presente documento no constituye un documento de política u orientación de la UPOV

1. El presente documento tiene por finalidad presentar un borrador de una nueva sección del documento TGP/8, Parte I: Diseño de ensayos DHE y análisis de datos sobre “Minimizar la variación resultante de la ejecución de los ensayos por distintos observadores”.

2. En el presente documento se utilizan las siguientes abreviaturas:

CAJ:	Comité Administrativo y Jurídico
TC:	Comité Técnico
TC-EDC:	Comité de Redacción Ampliado
TWA:	Grupo de Trabajo Técnico sobre Plantas Agrícolas
TWV:	Grupo de Trabajo Técnico sobre Hortalizas
TWC:	Grupo de Trabajo Técnico sobre Automatización y Programas Informáticos
TWF:	Grupo de Trabajo Técnico sobre Plantas Frutales
TWO:	Grupo de Trabajo Técnico sobre Plantas Ornamentales y Cultivos Forestales
TWP:	Grupos de Trabajo Técnico

3. La estructura del presente documento es la siguiente:

ANTECEDENTES.....	1
NOVEDADES ACAECIDAS EN 2014	2
COMITÉ TÉCNICO.....	2
GRUPOS DE TRABAJO TÉCNICO	2

ANEXO: PROYECTO DE ORIENTACIÓN PARA UNA FUTURA REVISIÓN DEL DOCUMENTO TGP/8 SOBRE LA MANERA DE MINIMIZAR LA VARIACIÓN RESULTANTE DE LA EJECUCIÓN DE LOS ENSAYOS POR DISTINTOS OBSERVADORES

ANTECEDENTES

4. En el documento TGP/8/1 Draft 7 PARTE I, párrafo 2.9.1: “Control of variation due to different observers” (Control de la variación resultante de la ejecución de los ensayos por distintos observadores), sometido al examen de los Grupos de Trabajo Técnico, en sus reuniones de 2007, se dispone:

“[Si esta sección fuera necesaria, se invita a los TWP a que contribuyan con orientación sobre el control de la variación resultante de la ejecución de los ensayos por distintos observadores cuando no se utiliza el análisis estadístico para determinar la distinción y a analizarla en relación con el párrafo 2.7.2.9.]”

5. Las novedades acaecidas antes de 2014 se presentan en el documento TC/50/21 “Revisión del documento TGP/8: Parte I: Diseño de los ensayos DHE y análisis de datos, nueva sección: Minimizar la variación resultante de la ejecución de los ensayos por distintos observadores”.

NOVEDADES ACAECIDAS EN 2014

Comité Técnico

6. En su quincuagésima sesión, celebrada en Ginebra del 7 al 9 de abril de 2014, el TC examinó el documento TC/50/21 “Revisión del documento TGP/8: Parte I: Diseño de los ensayos DHE y análisis de datos, nueva sección: Minimizar la variación resultante de la ejecución de los ensayos por distintos observadores”.

7. El TC tomó nota de que el TWF había solicitado que un experto de Nueva Zelanda informara, en su sesión de 2014, sobre la labor previa realizada en relación con la descripción armonizada de variedades de manzano para un conjunto de variedades acordado (véase el párrafo 46 del documento TC/50/36 “Informe sobre las conclusiones”).

8. El TC invitó al experto de Australia a que, con la ayuda de expertos de Alemania, Francia, los Países Bajos, el Reino Unido y la Unión Europea, elabore orientación adicional, para incluir en una futura revisión del documento TGP/8, sobre la manera de minimizar la variación resultante de la ejecución de los ensayos por distintos observadores, que incluya orientación relativa a los caracteres PQ y QN/MG, para someterla al examen de los TWP en sus sesiones de 2014 (véase el párrafo 47 del documento TC/50/36 “Informe sobre las conclusiones”).

9. En respuesta a la solicitud del TC, el redactor de Australia (Sr. Nik Hulse), con la ayuda de expertos de Alemania, Francia, los Países Bajos, el Reino Unido y la Unión Europea, elaboró un proyecto de orientación, para incluir en una futura revisión del documento TGP/8, sobre la manera de minimizar la variación resultante de la ejecución de los ensayos por distintos observadores, que incluye orientación relativa a los caracteres PQ y QN/MG, para someterla al examen de los TWP en sus sesiones de 2014.

10. El experto de Australia propuso que los TWP examinen las siguientes cuestiones para su aclaración o futura inclusión en el próximo proyecto de orientación, para una futura revisión del documento TGP/8, sobre la manera de minimizar la variación resultante de la ejecución de los ensayos por distintos observadores:

- Como ya se indica en el proyecto, los caracteres QN/MG pueden abordarse de modo similar a los QN/MS. No obstante, conviene señalar que, en el caso de los caracteres QN/MG, también debe tenerse en cuenta el posible efecto de la variación intraparceldaria aleatoria.
- Ha de estudiarse con más detenimiento el modo de incorporar la orientación relativa a los caracteres PQ. Debe explicarse que el documento TGP/14 es otro instrumento útil para aclarar las particularidades de muchos caracteres PQ (por ejemplo, la forma). Es posible que las diferencias entre los observadores puedan analizarse mediante la frecuencia de las desviaciones.
- Actualmente, el documento se centra en la variación entre observadores a nivel de la autoridad. Cabría considerar si la manera de minimizar la variación entre autoridades a nivel de los observadores debería abordarse en este documento o bien en un futuro documento aparte. Ha de tenerse en cuenta que con ello se introducirían más factores, tales como los métodos de muestreo, los *ring tests* y la variación debida a la interacción entre el genotipo y el entorno.

Grupos de Trabajo Técnico

11. El TWO, el TWF, el TWC, el TWV y el TWA examinaron respectivamente los documentos TWO/47/15, TWF/45/15, TWC/32/15, TWV/48/15 y TWA/43/15 “Revision of Document TGP/8: Part I: DUS Trial Design and Data Analysis, New Section: Minimizing the Variation due to Different Observers” (“Revisión

del documento TGP/8: Parte I: Diseño de los ensayos DHE y análisis de datos, nueva sección: Minimizar la variación resultante de la ejecución de los ensayos por distintos observadores”) (véanse los párrafos 39 a 41 del documento TWO/47/28 “Report”, los párrafos 26 a 32 del documento TWF/45/32 “Report”, los párrafos 50 a 53 del documento TWC/32/28 “Report”, los párrafos 34 a 37 del documento TWV/48/43 “Report” y los párrafos 28 a 33 del documento TWA/43/27 “Report”).

12. El TWO tomó nota de que el TWF había solicitado que un experto de Nueva Zelanda informara, en su sesión de 2014, sobre la labor previa realizada en relación con la descripción armonizada de variedades de manzano para un conjunto de variedades acordado, como se expone en el párrafo 18 del documento TWO/47/15 (véase el párrafo 40 del documento TWO/47/28 “Report”).

13. El TWO, el TWF, el TWC, el TWV y el TWA convinieron en que el proyecto de orientación presentado en sus sesiones como Anexo de los documentos TWO/47/15, TWF/45/15, TWC/32/15, TWV/48/15 y TWA/43/15, respectivamente, debe seguir elaborándose para su inclusión en una futura revisión del documento TGP/8 sobre la manera de minimizar la variación resultante de la ejecución de los ensayos por distintos observadores, que incluya orientación relativa a los caracteres PQ y QN/MG, junto con las cuestiones planteadas por el experto de Australia en los documentos TWO/47/15, TWF/45/15, TWC/32/15 y TWV/48/15, que se reproducen en el párrafo 10 del presente documento (véase el párrafo 41 del documento TWO/47/28 “Report”, el párrafo 27 del documento TWF/45/32 “Report”, el párrafo 52 del documento TWC/32/28 “Report”, el párrafo 34 del documento TWV/48/43 “Report” y el párrafo 30 del documento TWA/43/27 “Report”).

14. El TWO y el TWA convinieron en que el documento debe centrarse en la variación entre observadores a nivel de la autoridad y no en la manera de minimizar la variación entre autoridades a nivel de los observadores (véase el párrafo 41 del documento TWO/47/28 “Report” y el párrafo 31 del documento TWA/43/27 “Report”).

15. El TWF escuchó una ponencia a cargo de los expertos de Alemania y Nueva Zelanda sobre la labor previa realizada en relación con la descripción armonizada de variedades de manzano para un conjunto de variedades acordado, que se reproduce en el documento TWF/45/28 (véase el párrafo 28 del documento TWF/45/32 “Report”).

16. El TWF convino en la importancia de minimizar la variación entre distintos observadores así como entre autoridades, y por ello propuso considerar la posibilidad de emprender un nuevo proyecto sobre la descripción armonizada de variedades para un conjunto de variedades acordado. El experto de Alemania propuso presentar al TWF, en su cuadragésima sexta sesión, un protocolo para el proyecto que incluya una lista acordada de las variedades que han de examinarse, a fin de analizar si es pertinente seguir desarrollando el estudio (véase el párrafo 30 del documento TWF/45/32 “Report”).

17. El TWF y el TWV señalaron la importancia de la calidad de las directrices de examen como orientación clara para los encargados de realizar el examen DHE y como garantía de la consistencia de las observaciones efectuadas por distintos observadores en cada autoridad, así como la importancia de la formación continua de los examinadores (véase el párrafo 31 del documento TWF/45/32 “Report” y el párrafo 35 del documento TWV/48/43 “Report”).

18. El TWF invitó al experto de Australia a que, en su cuadragésima sexta sesión, presente información sobre el efecto del lugar, el observador y el año en la conformidad de un carácter para un cultivo determinado (véase el párrafo 32 del documento TWF/45/32 “Report”).

19. El TWC tomó nota de que, en la cuadragésima quinta sesión del TWF, el experto de Nueva Zelanda había presentado información sobre la labor previa realizada en relación con la descripción armonizada de variedades de manzano para un conjunto de variedades acordado (véase el párrafo 51 del documento TWC/32/28 “Report”).

20. El TWC convino en que el proyecto de orientación debe centrarse inicialmente en la variación entre observadores a nivel de la autoridad y, en una etapa ulterior, abordar la variación entre autoridades (véase el párrafo 53 del documento TWC/32/28 “Report”).

21. El TWV propuso incluir, en un ensayo DHE, un ejercicio de formación que sirva de base para intercambiar experiencias prácticas y para potenciar la utilización del TWV con fines de capacitación (véase el párrafo 36 del documento TWV/48/43 “Report”).

22. El TWA tomó nota de que, en la cuadragésima quinta sesión del TWF, un experto de Nueva Zelanda había presentado una ponencia sobre la labor previa realizada en relación con la descripción armonizada de variedades de manzano para un conjunto de variedades acordado, que se reproduce en el documento TWF/45/28 "*Harmonized example varieties for Apple: historical data and possible new developments*" ("Variedades ejemplo armonizadas de la manzana: datos históricos y posibles novedades") (véase el párrafo 29 del documento TWA/43/27 "*Report*").

23. El TWA señaló la importancia de la calidad de las directrices de examen como orientación clara para los encargados de realizar el examen DHE y como garantía de la consistencia de las observaciones, así como la importancia de la formación continua de los examinadores. El TWA acordó proponer como recomendación general que, si es posible, se asigne un solo observador a cada ensayo para evitar la variación en las observaciones (véase el párrafo 32 del documento TWA/43/27 "*Report*").

24. El TWA convino en que los caracteres QN/MG pueden abordarse de modo similar a los QN/MS y señaló que también debe tenerse en cuenta el posible efecto de la variación intraparcular aleatoria. Asimismo, convino en que las diferencias entre los observadores respecto de los caracteres PQ podrían analizarse mediante métodos no paramétricos, como la frecuencia de las desviaciones (véase el párrafo 33 del documento TWA/43/27 "*Report*").

25. En respuesta a las observaciones formuladas por los TWP en sus sesiones de 2014, el Sr. Nik Hulse (Australia) ha elaborado un nuevo proyecto de orientación para su inclusión en el documento TGP/8, Parte I: Diseño de los ensayos DHE y análisis de datos, nueva sección: Minimizar la variación resultante de la ejecución de los ensayos por distintos observadores, que se reproduce en el Anexo del presente documento, para que los TWP lo examinen en sus sesiones de 2015.

26. Se invita al TC a:

a) *examinar el proyecto de orientación, para su inclusión en el documento TGP/8, sobre la manera de minimizar la variación resultante de la ejecución de los ensayos por distintos observadores, que se reproduce en el Anexo del presente documento, junto con las observaciones formuladas por los TWP en sus sesiones de 2014; y*

b) *solicitar a los TWP que, en sus sesiones de 2015, examinen el nuevo proyecto, que incorporará las eventuales modificaciones que acuerde el TC en su quincuagésima primera sesión.*

[Sigue el Anexo]

TGP/8/1: PARTE I: NUEVA SECCIÓN: MINIMIZAR LA VARIACIÓN RESULTANTE DE LA EJECUCIÓN DE UN ENSAYO POR DISTINTOS OBSERVADORES

1. Introducción

El objeto del presente documento lo constituyen los caracteres QN/MG, QN/MS, QN/VG y QN/VS. En él no se abordan de manera explícita caracteres PQ como el color y la forma. El método kappa que aquí se describe es, en sí mismo, aplicable en gran medida a estos caracteres; por ejemplo, el coeficiente estándar kappa está concebido para datos nominales. No obstante, hasta donde sabemos, este método no se ha utilizado con caracteres PQ y, además, para la calibración de dichos caracteres puede ser necesaria información adicional. A modo de ejemplo, para la calibración del color también ha de tenerse en cuenta la carta de colores RHS, las condiciones de iluminación, etcétera. Estos aspectos no se incluyen en el presente documento.

1.1 La variación en las mediciones o las observaciones puede estar causada por distintos factores, como el tipo de cultivo o de carácter, el año, el lugar, el diseño y la gestión de los ensayos, el método y el observador. En lo que atañe, en particular, a los caracteres evaluados visualmente (QN/VG o QN/VS), es posible que las diferencias entre los observadores den lugar a una gran variación y una eventual distorsión en las observaciones. Es posible que un observador esté menos formado que otro o interprete el carácter de manera distinta. Así pues, si el observador A evalúa la variedad 1 y el observador B mide la variedad 2, la diferencia observada podría deberse a las diferencias entre los observadores A y B, en lugar de tratarse de diferencias entre las variedades 1 y 2. Claramente, lo que más interesa son las diferencias entre las variedades y no entre los observadores. Cabe señalar que si bien la variación causada por distintos observadores no puede eliminarse, sí puede controlarse.

1.2 En la medida de lo posible, se recomienda asignar un solo observador a cada ensayo para minimizar la variación en las observaciones resultante de la ejecución de los ensayos por distintos observadores.

2. Capacitación e importancia de explicar con claridad los caracteres y el método de observación

2.1 La capacitación de los nuevos observadores es fundamental para la coherencia y la continuidad de las observaciones de variedades vegetales. Recursos útiles en este sentido son los manuales de calibración y la supervisión y orientación a cargo de observadores expertos, así como el empleo de variedades ejemplo que ilustren la gama de expresión.

2.2 Con las directrices de examen de la UPOV se procura armonizar el proceso de descripción de las variedades y describir lo más claramente posible los caracteres de un cultivo y los niveles de expresión. Se trata del primer paso para controlar la variación y la distorsión. Sin embargo, la forma de observar o medir un carácter podrá variar en función del lugar o de la autoridad encargada del examen. Los manuales de calibración realizados por las autoridades locales de examen son muy útiles para aplicar las directrices de examen de la UPOV a escala local. En caso necesario, esos manuales, específicos para los distintos cultivos, explican los caracteres que hay que observar en mayor detalle y especifican cuándo y cómo deberían ser observados. Además, es posible que contengan fotografías y dibujos de cada carácter, a menudo para cada nivel de expresión de un carácter.

2.4 El Glosario de términos utilizados en los documentos de la UPOV (TGP/14/2) constituye una orientación útil para aclarar las particularidades de muchos caracteres, en especial los caracteres PQ.

2.3 Una vez completada la capacitación, es importante velar por que, con cierta frecuencia, el observador se someta a recalibración y reciba formación de actualización.

3. Prueba de la calibración

3.1 El paso siguiente a la capacitación del observador podría ser probar el desempeño de éste en un experimento de calibración. Ello es útil en particular para los observadores inexpertos que tengan que realizar observaciones visuales (caracteres QN/VG y QN/VS). Si realizan observaciones visuales, de preferencia, deberían superar un examen de calibración antes de realizar observaciones en el ensayo. Sin embargo, también para los observadores expertos es útil ponerse a prueba periódicamente para verificar que aún satisfacen los criterios de calibración.

3.2 Puede prepararse un experimento de calibración que se analizará de maneras distintas; por lo general, supone múltiples observadores que miden el mismo material, tras lo cual se evalúan las diferencias entre los observadores.

4. Prueba de la calibración para los caracteres QN/MG o QN/MS

4.1 Para las observaciones realizadas con instrumentos de medición, como reglas (a menudo, los caracteres QN/MS), la medición suele realizarse en escala de intervalo o de relación. En este caso, puede aplicarse el enfoque de Bland y Altman (1986), que comienza con un diagrama de los resultados tomados por un par de observadores en un diagrama de dispersión, que se compara con la línea de igualdad (en la cual $y=x$). Esto ayuda a evaluar a ojo el grado de concordancia entre mediciones del mismo objeto. El paso siguiente es tomar la diferencia por objeto y construir un gráfico en el que el eje "y" represente la diferencia entre los observadores, y el eje "x" o bien el índice del objeto, o bien el valor medio del objeto. Si además se dibujan las líneas horizontales $y=0$, $y = \text{media}(\text{diferencia})$ y las dos líneas $y = \text{media}(\text{diferencia}) \pm 2 \times \text{desviación estándar}$, puede encontrarse fácilmente, por un lado, la distorsión entre los observadores y, por el otro, cualquier valor atípico. Del mismo modo se puede evaluar también la diferencia entre la medición de cada observador y el promedio de las mediciones de todos los observadores. Pueden aplicarse métodos de prueba, como la prueba de la t por pares, para comprobar si existe una desviación significativa entre un observador y otro o la media de los demás observadores.

4.2 Si se toman dos mediciones de cada objeto efectuadas por un mismo observador, cabe examinar las diferencias entre ambas. Si dichas diferencias son grandes en comparación con las obtenidas por otros observadores, es posible que este observador presente una baja repetibilidad. A partir del recuento de los valores atípicos moderados y grandes obtenidos por cada observador (por ejemplo, valores mayores que dos veces y tres veces la desviación típica, respectivamente), puede elaborarse un cuadro que confronte observadores con número de valores atípicos, que puede utilizarse para decidir si un observador satisface los criterios de garantía de calidad.

4.3 Pueden realizarse otros controles de calidad basados en las pruebas de repetibilidad y reproducibilidad para laboratorios que se describen en la norma ISO 5725-2. En el sitio web de la ISTA se ofrecen programas informáticos gratuitos que permiten obtener valores y gráficos con arreglo a esta norma ISO.

4.4 En muchos casos de QN/MG o QN/MS, bastará por lo general con instrucciones correctas y claras para que la variación o la distorsión entre los observadores en las mediciones sea a menudo insignificante. Si hay lugar a duda, un experimento de calibración como el que se describió *supra* puede ayudar a poner en claro la situación.

4.5 Cuando se observen caracteres QN/MG, se deberá tener en cuenta la posible variación intraparcelar.

5. Prueba de la calibración para los caracteres QN/VS o QN/VG

5.1 Para el análisis de los datos ordinales (caracteres QN/VS o QN/VG), puede resultar muy instructiva la creación de cuadros de contingencia de los distintos resultados obtenidos por cada par de observadores. Para comprobar la diferencia estructural (distorsión) entre dos observadores, puede utilizarse la prueba de pares igualados de Wilcoxon (denominada también prueba de Wilcoxon de rangos señalados).

5.2 Para medir el grado de concordancia, suele utilizarse el coeficiente kappa de Cohen (Cohen, 1960), destinado a encontrar la concordancia aleatoria: $\kappa = P(\text{concordancia}) - P(e) / (1 - P(e))$, siendo $P(\text{concordancia})$ la fracción de objetos que están en la misma clase para ambos observadores (la diagonal principal en el cuadro de contingencia), y $P(e)$ la probabilidad de concordancia aleatoria, teniendo en cuenta los marginales (como en una prueba de ji cuadrado). Si existe concordancia perfecta entre los observadores, el valor kappa es $\kappa = 1$. Si no hay concordancia entre los observadores, aparte de lo previsto debido al azar ($P(e)$), entonces $\kappa = 0$.

5.3 El coeficiente estándar kappa de Cohen toma en consideración únicamente la concordancia perfecta contra la no concordancia. Si se desea tener en cuenta el grado de concordancia (por ejemplo, con los caracteres ordinales), puede aplicarse un coeficiente kappa lineal o ponderado cuadrático (Cohen, 1968). Si se desea tener una única estadística simultáneamente para todos los observadores, puede calcularse un coeficiente kappa generalizado. La mayoría de los programas informáticos de cálculo estadístico, entre otros, SPSS, Genstat y R (el programa Concord), incluye instrumentos para calcular el coeficiente kappa.

5.4 Como se ha observado, un valor κ bajo indica escasa concordancia, y valores cercanos a 1 indican excelente concordancia. Suele considerarse que los valores entre 0,6 y 0,8 indican una concordancia

importante y que los valores por encima de 0,8 indican una concordancia prácticamente perfecta. De ser necesario, pueden utilizarse valores z para kappa (suponiendo que haya una distribución aproximadamente normal). A los expertos en DHE con experiencia cabría aplicarles criterios más estrictos que a los observadores inexpertos.

6. Diseño de los ensayos

6.1 Si hay varios observadores en un ensayo, lo mejor es que una persona observe una o más repeticiones completas. En ese caso, la corrección de los efectos de bloque también tiene en cuenta la distorsión entre los observadores. Si es necesario más de un observador por repetición, deberá prestarse particular atención a la calibración y la concordancia. En algunos casos, puede ser útil valerse de diseños de bloques incompletos (como los diseños alfa), y puede asignarse un observador a los sub-bloques. De esta forma pueden corregirse las diferencias sistemáticas entre los observadores.

7. Ejemplo del coeficiente kappa de Cohen

7.1 En este ejemplo hay 3 observadores y 30 objetos (parcelas o variedades). El carácter se observa en una escala de 1 a 6. Los datos en bruto y sus valores figuran en los cuadros siguientes.

Variedad	Observador 1	Observador 2	Observador 3
V1	1	1	1
V2	2	1	2
V3	2	2	2
V4	2	1	2
V5	2	1	2
V6	2	1	2
V7	2	2	2
V8	2	1	2
V9	2	1	2
V10	3	1	3
V11	3	1	3
V12	3	2	2
V13	4	5	4
V14	2	1	1
V15	2	1	2
V16	2	2	3
V17	5	4	5
V18	2	2	3
V19	1	1	1
V20	2	2	2
V21	2	1	2
V22	1	1	1
V23	6	3	6
V24	5	6	6
V25	2	1	2
V26	6	6	6
V27	2	6	2
V28	5	6	5
V29	6	6	5
V30	4	4	4

Resultados por variedad	1	2	3	4	5	6
V1	3	0	0	0	0	0
V2	1	2	0	0	0	0
V3	0	3	0	0	0	0
V4	1	2	0	0	0	0
V5	1	2	0	0	0	0
V6	1	2	0	0	0	0
V7	0	3	0	0	0	0
V8	1	2	0	0	0	0
V9	1	2	0	0	0	0
V10	1	0	2	0	0	0
V11	1	0	2	0	0	0
V12	0	2	1	0	0	0
V13	0	0	0	2	1	0
V14	2	1	0	0	0	0
V15	1	2	0	0	0	0
V16	0	2	1	0	0	0
V17	0	0	0	1	2	0
V18	0	2	1	0	0	0
V19	3	0	0	0	0	0
V20	0	3	0	0	0	0
V21	1	2	0	0	0	0
V22	3	0	0	0	0	0
V23	0	0	1	0	0	2
V24	0	0	0	0	1	2
V25	1	2	0	0	0	0
V26	0	0	0	0	0	3
V27	0	2	0	0	0	1
V28	0	0	0	0	2	1
V29	0	0	0	0	1	2
V30	0	0	0	3	0	0

El cuadro de contingencia para los observadores 1 y 2 es:

O1\O2	1	2	3	4	5	6	Total
1	3	0	0	0	0	0	3
2	10	5	0	1	0	1	17
3	2	1	0	0	0	0	3
4	0	0	0	1	0	0	1
5	0	0	0	1	0	2	3
6	0	0	1	0	0	2	3
Total	15	6	1	3	0	5	30

El coeficiente kappa entre el observador 1 y el observador 2, $\kappa(O1,O2)$, se calcula de la manera siguiente:

- $\kappa(O1,O2) = P(\text{concordancia entre O1 y O2}) - P(e) / (1 - P(e))$, siendo:
- $P(\text{concordancia}) = (3+5+0+1+0+2)/30 = 11/30 \approx 0,3667$ (elementos diagonales)
- $P(e) = (3/30).(15/30) + (17/30).(6/30) + (3/30).(1/30) + (1/30).(3/30) + (3/30).(0/30) + (3/30).(5/30) \approx 0.1867$ (márgenes por pares)
- Es decir que: $\kappa(O1,O2) \approx (0.3667-0.1867) / (1-0.1867) \approx 0,22$

Se trata de un valor bajo, que indica una concordancia muy leve entre esos dos observadores; en ese caso la situación sería preocupante y cabría tomar medidas para mejorar la concordancia. De manera similar, pueden calcularse los valores para otros pares: $\kappa(O1,O3) \approx 0.72$, $\kappa(O2,O3) \approx 0.22$. Se constata una buena concordancia entre el observador 1 y el observador 3. El observador 2 presenta una clara diferencia respecto del observador 1 y el observador 3 y es preciso investigar los motivos de tal desviación (debe considerarse, por ejemplo, si necesita capacitación adicional).

8. Referencias

Cohen, J. (1960) "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement* 20: 37-46.

Cohen, J. (1968) "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit". *Psychological Bulletin*, 70(4): 213-220.

Bland, J. M. Altman D. G. (1986) "Statistical methods for assessing agreement between two methods of clinical measurement", *Lancet*: 307-310.

<http://www.seedtest.org/en/stats-tool-box-content--1--1143.html> (programas informáticos basados en la norma ISO 5725-2)

[Fin del anexo y del documento]