

INTERPRETACIÓN DE LA DISPERSIÓN DE DATOS EN CONTEXTO DE RIESGO POR ESTUDIANTES DE SECUNDARIA^{xxxii}

Interpreting spread of data in risk context by middle school students

José Antonio Orta y Ernesto Sánchez

Departamento de Matemática Educativa, Cinvestav-IPN, México

Resumen

La investigación tiene como objetivo explorar el razonamiento de los estudiantes acerca de la noción de dispersión (variabilidad o variación) cuando analizan datos en situaciones de incertidumbre. En particular, en esta comunicación se informa sobre las respuestas a dos problemas de un cuestionario administrado a 65 estudiantes de 9º grado (14 años). Los problemas son de comparación de conjuntos de datos y están ubicados en contextos de riesgo: apuestas en juegos y duración de vida después de tratamientos médicos. El cuestionario fue aplicado antes y después de unas actividades de enseñanza. Los resultados muestran la dificultad de los estudiantes para interpretar la dispersión en contextos de riesgo. Aunque identifican el conjunto de datos con mayor dispersión, no es suficiente para que la interpreten y tomen una decisión racional.

Palabras clave: *dispersión, variabilidad, incertidumbre, riesgo.*

Abstract

The aim of this investigation is to explore the student reasoning about variation (variability or spread) when they analyze data in situations under uncertainty. In particular, in this communication the responses to two problems of a questionnaire administered to 65 ninth-grade students (14 years old) are reported. The problems are of comparing groups of data in situations of risk: stakes in games and the life expected after medical treatments. Students responded the questionnaire before and after instruction activities. The results show the difficulty found by students to interpret variation in risk context. Although they indentify the data set with more variation, it is not enough for interpreting it in the risk context and makes a rational decision.

Keywords: *spread, variability, uncertainty, risk*

INTRODUCCIÓN

Muchos autores han destacado la importancia de la variación estadística. Moore (1990) afirma que es el corazón de la estadística, mientras que Watson, Kelly, Callingham y Shaughnessy (2003) la ubican como la razón de existir de la estadística. Wild y Pfannkuch (1999) proponen que la consideración de la variación es uno de los tipos fundamentales del pensamiento estadístico. Este concepto está relacionado con varias ideas estadísticas fundamentales: representación, centros, distribución, inferencia. Garfield y Ben-Zvi (2008) observan que “la comprensión de las ideas de dispersión o variación en los datos es una componente clave en la comprensión del concepto de distribución y es esencial para hacer inferencias estadísticas” (p. 203). A pesar de su importancia, la variación estadística en los niveles básicos está prácticamente ausente. En el currículo mexicano, la estadística de secundaria se reduce principalmente a contenidos de gráficas y medidas de tendencia central y sólo al final del tercer grado (14 años) se menciona el estudio de “medidas de dispersión”, sugiriendo tratar el rango y la desviación media (SEP, 2011).

En los estudios arriba mencionados y en muchos otros sobre variación estadística (Shaughnessy, 2007, Sánchez, da Silva y Coutinho, 2011) se sugiere que el estudio de la dispersión no puede reducirse al aprendizaje de las fórmulas correspondientes: rango, desviación media, desviación estándar o y/o variancia. Es necesario ampliar el conjunto de ideas alrededor de la dispersión, en particular, son fundamentales sus interpretaciones en los diferentes contextos en los que se presenta. Por otra parte, tampoco se puede analizar la comprensión de la dispersión aislada de la idea de centro; por ejemplo, Konold y Pollatsek (2004) utilizan la metáfora de señal en presencia de ruido para explicar la relación entre centro y dispersión; Garfield y Ben-Zvi (2008, p. 203) afirman que “es imposible considerar la variabilidad sin considerar también el centro, pues ambas ideas son necesarias para encontrarles significado en el análisis de datos”. La presente comunicación tiene como objetivo dar cuenta de cómo interpretan los estudiantes de 9º grado la dispersión de los datos en situaciones de riesgo. Aunque sin ser el objeto de este estudio, en la descripción y el análisis se hace también consideraciones sobre la media aritmética. .

Conviene aclarar que utilizaremos los términos *variación*, *variabilidad* y *dispersión* como sinónimos, apoyándonos en el análisis de Estepa (2013). Aunque estrictamente hablando tales términos conllevan ligeras diferencias, en el contexto presente, su uso se refiere al núcleo que comparten.

MARCO DE REFERENCIA

La interpretación de la dispersión depende de la situación de la cual provienen los datos, en particular, se pueden distinguir dos interpretaciones algo generales sugeridas por el tipo de contexto del problema y la clase de distribución a la que pertenecen los datos. Una es la dispersión como una medida de la exactitud y precisión de un instrumento de medida o un proceso de medición; en estos casos los datos pertenecen a una distribución normal. La otra, como una medida de la incertidumbre presente en el resultado de un proceso; si esta incertidumbre implica alguna amenaza en la bondad de un resultado, entonces se llama riesgo; las clases de distribuciones a la que pertenecen suelen ser muy alejadas de una forma normal.

El riesgo se presenta cuando hay potenciales resultados no deseados que pueden traer como consecuencia pérdidas o daños. Definir el riesgo significa especificar los resultados valiosos y los no deseados en un orden que refleje el valor que se les atribuye. El análisis del riesgo ofrece información para la toma de decisiones. La teoría de la toma de decisiones en situaciones de riesgo tiene dos aspectos; por un lado, define reglas abstractas sobre lo que debería hacer la gente, y por otro lado, estudia lo que hace realmente cuando se enfrenta a ellas: “Si la gente no sigue las reglas,

quiere decir que o las personas necesitan ayuda o las reglas necesitan una revisión” (Fischhoff y Kadavy, 2011, pp.65).

En los estudios descriptivos sobre la toma de decisiones, se han identificado dos actitudes de los sujetos frente a situaciones de riesgo: propensión y aversión al riesgo (Tversky y Kahneman, 2000). Una situación puede ser enfocada desde un punto de vista conservador o aventurero; así ante dos situaciones de juego en las que se gana en promedio lo mismo, pero en un juego con apuestas pequeñas y en el otro con apuestas grandes, alguien conservador preferirá el primero y alguien aventurero preferirá el segundo.

METODOLOGÍA

Este es un estudio de carácter exploratorio sobre la percepción e interpretación espontánea de los estudiantes frente a problemas en situaciones de riesgo. Se trata de revelar el tipo de razonamiento que despliegan los estudiantes en situaciones de riesgo. Los contextos son dos, uno de apuestas en juegos de azar, el otro de tratamientos médicos. Esta elección se hizo con el objetivo de atraer el interés de los estudiantes hacia la dispersión.

Participantes. Participaron 65 estudiantes, un profesor y los autores. Los estudiantes estaban distribuidos en dos grupos que cursaban su tercer año de secundaria (9º grado, 14 años) en una escuela pública de la Ciudad de México. El profesor tiene 12 años de experiencia y realizó una maestría en educación matemática.

Instrumentos. Se diseñaron y llevaron a cabo dos actividades de enseñanza y un cuestionario de tres problemas que se administraron antes y después de las actividades de enseñanza. En esta comunicación se informa de los resultados de dos problemas del cuestionario (ver anexo). Estos se diseñaron siguiendo el formato de comparación de grupos, ya que esta clase de problemas se ha mostrado fértil para poner en acción el razonamiento de los estudiantes (Garfield y Ben-Zvi, 2008). Cada problema tenía dos incisos, uno en una situación de toma de decisiones; en el problema 1 tenían que elegir el juego en el que les convendría jugar (más ganancia); en el problema 2, el tratamiento que recomendarían (mayor tiempo de vida). En ambos problemas las medias son las mismas de manera que las diferencias se buscaran interpretando la dispersión.

Las actividades de enseñanza tuvieron como eje dos problemas uno en contexto de medición y otro en contexto de juego; ambos contextos no implicaban riesgo. Los problemas también fueron de comparación de conjuntos de datos y se condujo a los estudiantes a considerar el rango y la desviación media. Estas actividades se llevaron a cabo entre las dos aplicaciones del cuestionario (previa y posterior).

Procedimiento. El cuestionario se aplicó a los estudiantes antes de realizar la actividad (Previo), antes ya habían estudiado el tema de medidas de tendencia central, de manera que sabían calcular la media y mediana de un conjunto de datos. A lo largo de 4 sesiones de 100 minutos, se realizaron las actividades. Después, en una quinta sesión, se les volvió a aplicar el cuestionario (Posterior). Se esperaba que mejorara su desempeño en la segunda aplicación como resultado de la actividad.

RESULTADOS

A continuación se describen los resultados de dos problemas del cuestionario que se administró antes y después de la intervención.

Situación 1. Esta se refiere a las cantidades ganadas por diferentes personas, en dos juegos (Ver anexo) Las preguntas son 1a) ¿en cuál de los dos juegos participarías? y 1b) ¿en cuál de los juegos hay más variabilidad? Se esperaba que los estudiantes percibieran que en ambos juegos se gana en

promedio lo mismo, y que esto llevara a algunos a responder que “cualquiera”; otros considerarían la dispersión y quienes tuvieran aversión al riesgo elegirían juego 2, mientras quienes fueran propensos al riesgo, el juego 1. En la Tabla 1, se indican las frecuencias de elección en las respuestas de los estudiantes.

Tabla 1. Frecuencias con las que los estudiantes eligieron como respuesta a las preguntas 1a y 1b, “Juego 1”, “Juego 2”, “cualquiera” en el cuestionario previo y en el posterior.

	Pregunta 1a)		Pregunta 1b)	
	Previo	Posterior	Previo	Posterior
Juego 1	29	25	44	39
Juego 2	33	33	20	24
Cualquiera	3	8	1	4
No responde	0	1	0	0
Total	65	67	65	67

Conviene observar que los datos de ambos juegos tienen la misma media y que el juego 1 tiene mayor dispersión que el juego 2. Con relación a la pregunta 1a, en el Posterior hay 4 respuestas menos que eligen el juego 1 (mayor variación) y aumentan las respuestas que proponen elegir cualquier de los dos. Este leve cambio podría explicarse apelando a que se repasó la media durante la intervención, pues aumentó el número de estudiantes que respondieron “cualquiera”. En el inciso b, se observa que en el Posterior, hay 5 elecciones menos al juego 1 que al juego 2; aumenta en 4 los que eligen el juego 2 y sólo aumenta en 3 los que opinan que cualquier juego se puede elegir.

Se nota que en la pregunta 1b hay más elecciones del juego 1 (tanto en el previo como en el posterior) respecto a las elecciones del mismo juego en el inciso a. Esto quiere decir que varios identifican el juego en que los datos tienen más variabilidad, pero que en el contexto de la apuesta eligen la opción del juego con menos variabilidad o la opción de que cualquiera es igual. Sin embargo, tal identificación no es muy sólida, como se ve enseguida.

Es notable que en esta pregunta (1b), después de la intervención, haya disminuido el número de respuestas correctas; por lo que conviene examinar con más detalle los cambios que del pre al post se realizaron en esta pregunta. En la tabla dos se muestran tales cambios.

Tabla 2. Frecuencias de estudiantes que cambiaron su elección entre juego 1 y 2

	Cuestionario previo	Cuestionario Posterior		
		Juego 1	Juego 2	Juego 1 y Juego 2
		Juego 1	25	14
Juego 2	11	6	2	
Juego 1 y Juego 2	1	0	0	

* La suma de esta fila es 41 y no 44 (tabla 1) porque sólo 61 resolvieron ambos, el pre y el post.

Lo más significativo es que 14 estudiantes que eligieron en el pre el juego 1 (respuesta correcta) en el posterior cambiaron al juego 2. Esto significa que la enseñanza no logró desarrollar y consolidar su primera intuición. En nuestra opinión, su primera respuesta fue motivada por su conocimiento común de lo extenso y lo disperso; después de la instrucción, buscaron mejores razones para su elección pero no supieron considerar los elementos correctos en los datos y esto los desvió. Pero es necesaria mayor evidencia para confirmar esta opinión.

Se esperaba que los estudiantes interpretaran la variabilidad y eligieran el juego 1, cuando son propensos al riesgo y el juego 2 cuando tienen aversión al riesgo, sin embargo, las justificaciones que dieron o no dan información sobre sus verdaderos motivos para elegir una u otra opción o indican motivos que se estiman inadecuados, más bien basados en la comparación de algún o algunos elementos particulares de los datos. A continuación se dan ejemplos que muestran lo anterior.

Con relación al inciso a, una gran parte de las “justificaciones” de los estudiantes para sostener sus decisiones no aportan información que permita revelar sus verdaderos motivos (30 en el previo y 23 en el posterior), por ejemplo: “Juego 1, porque se gana más”. Hay un número reducido que hacen comentarios que ofrecen indicios de sus posibles estrategias. Se clasifican las justificaciones en tres tipos:

1) Suman la totalidad de datos de cada juego y los comparan:

a) ¿En cuál de los dos juegos participaría? En cualquiera de los dos
 ¿Por qué? Por que los dos tienen la misma variabilidad de 12,000 en cada juego

“En cualquiera de los dos. Porque los dos tienen la misma variabilidad de 12,000 en cada juego” [Confunde ‘variabilidad’ con ‘cantidad total’]

2) Asumen que en el conjunto con mayor variación se gana más:

a) ¿En cuál de los dos juegos participaría? en el 1º
 ¿Por qué? porque podría ganar más dinero ya que hay más variabilidad

“En el 1, porque podría ganar más dinero ya que hay más variación”

3) Buscan en cada conjunto de datos, representantes que consideran que indican en qué juego se gana más (valores máximos, valores mínimos, mayor frecuencia) y los comparan.

a) ¿En cuál de los dos juegos participaría? En el juego 1
 ¿Por qué? por que al final la ganancia es mayor

“En el juego 1, porque al final la ganancia es mayor”

Con relación al inciso b, la mayoría identifica el conjunto de datos con mayor dispersión; no obstante, unos pocos utilizan en su justificación el rango y ninguno considera otra medida de dispersión (La desviación media se trabajó durante la intervención); por ejemplo:

b) ¿En cuál de los juegos hay más variabilidad? En el juego 1
 Explica tu respuesta por que en el juego 1 están más separados los números y en el juego 2 los números están más juntos me refiero en el juego 1 van desde el 870 hasta el 1573 y en el juego 2 van desde 1110 hasta 1335

“En el juego 1, porque en el juego 1 están más separados los números y en el juego 2 los números están más juntos. Me refiero [a que] en el juego 1 van desde el 870 hasta el 1573 y en el juego 2 van desde 1110 hasta 1335”.

Situación 3. Esta se refiere al tiempo de vida de pacientes enfermos que siguen diferentes tratamientos (ver apéndice). Las preguntas son 3a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? ¿Por qué? y 3b) ¿En cuál de los tratamientos hay más variación? Explica tu respuesta..

Tabla 3. Frecuencias con las que los estudiantes eligieron como respuesta a la pregunta 3a) y 3b), “Tratamiento 1”, “Tratamiento 2”, “Tratamiento 3”, etcétera, en el cuestionario previo y en el posterior.

	Pregunta 3a)		Pregunta 3b)	
	C Previo	C Posterior	C Previo	C Posterior
Tratamiento 1	30	30	42	51
Tratamiento 2	19	23	9	6
Tratamiento 3	15	12	9	6
Tratamientos 2 y 3	0	0	5	0
Cualquiera	1	1	0	0
No contestó	0	1	0	4
Total	65	67	65	67

Los datos de los tres tratamientos tienen la misma media; el tratamiento 1 tiene mayor dispersión que los otros y el tratamiento 2 tiene menor desviación media.

En la Tabla 3, se puede observar que, con relación a la pregunta 3a, la frecuencia con la que eligieron el tratamiento 1 en el previo es la misma que la frecuencia con la que lo eligieron en el posterior; mientras que el tratamiento 2 tuvo 4 elecciones más en el posterior respecto al previo; en cambio, la elección del tratamiento 3 disminuyó en 3 del previo al posterior. Los cambios son poco significativos; aparentemente la intervención no influyó mayormente en el razonamiento de los estudiantes para este tipo de pregunta.

En la parte 3b, hay 9 elecciones más del tratamiento 1 en el posterior que en el previo, lo que indica una mejoría en la identificación de la distribución con mayor dispersión. También hubo 3 elecciones menos de los tratamientos 2 y 3 al pasar del previo al posterior. Hubo cinco elecciones de los tratamientos 2 y 3 en el previo, pero en el posterior nadie dio esta respuesta.

Con relación a las justificaciones de las respuestas a la pregunta 3a, se encuentran argumentos de tres tipos 1) apoyan su elección con una afirmación general que de alguna manera repite lo que se pide (tautológica), 2) Suman la totalidad de tiempo que tardan los tratamientos (les falta dividir entre el número de datos para que obtengan la media), 3) Observan un valor particular de los datos de un tratamiento y lo comparan con el correspondiente de otro o de los otros tratamientos. A continuación se presentan un ejemplo de cada caso:

Tautológico

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? el 1
 ¿Por qué? Viven mas años

“El 1, viven más años)

Se fija en la suma de los tiempos de tratamiento

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? en el que sea
 ¿Por qué? sumando los años de tratamiento y los pacientes el resultado son los mismos

“Sumando los años de tratamiento y los pacientes, el resultado es el mismo”

Se fija en los mínimos y los compara

- a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? El tratamiento 2
 ¿Por qué? la esperanza de vida, aunque no es la más alta, el mínimo es mayor.

“La esperanza d vida, aunque no es la más alta, el mínimo es mayor”

CONCLUSIONES

Las evidencias apoyan la afirmación de que para los estudiantes de la edad de 14 años, es muy difícil que espontáneamente interpreten la dispersión en contextos de riesgo. Sus tendencias a comparar sólo valores particulares de los juegos (máximos, mínimos, o modas) o de los tratamientos, y tomar decisiones basados sólo en esa comparación, remiten a la dificultad, ya documentada, consistente en que los estudiantes no pueden hacer comparaciones de grupos de datos considerando cada grupo como una unidad o distribución (Garfield y Ben-Zvi, 2008). Al parecer, la intervención no ayudó a mejorar las interpretaciones de los estudiantes. En ésta, se estudiaron los temas de gráficas, media y dispersión en contextos diferentes a los de riesgo, y se constató que ahí alcanzaron un cierto nivel de comprensión, pero no fue suficiente para que tales conceptos fueran transferidos e interpretados en los contextos de riesgo presentes en las preguntas del cuestionario.

Las respuestas de los incisos b de las preguntas revelan que la mayoría de los estudiantes perciben cuál de dos conjuntos de datos tienen mayor dispersión, pero también revelan que tal percepción es muy primitiva, pues una gran parte de ellos no se apoya en alguna característica de los datos (por ejemplo, el rango) sino que en sus justificaciones sólo afirman de otra manera lo que se pregunta (“porque se gana más” o “porque se vive más”). Responder que hay más variabilidad o dispersión en un conjunto de datos está más motivado por sus conocimientos comunes sobre lo extenso o lo disperso que por un conocimiento estadístico. Aunque esta investigación tiene el objetivo de enfocarse en la dispersión no puede dejarse de lado considerar a la media. Es conveniente señalar que la mayoría de los estudiantes participantes no utiliza la media en sus valoraciones, por ejemplo, para concluir que la ganancia promedio es la misma en los dos juegos, así como que el tiempo promedio de vida es el mismo en ambos tratamientos. Se esperaba que el análisis de la dispersión que hicieran los estudiantes se articulara con una consideración de la media, lo cual no ocurrió sino en muy pocos casos.

Referencias

- Estepa, C. A. (2013). Los fenómenos de cambio. *I Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria*. Organizado por el grupo de investigación Didáctica de la Estadística, Probabilidad y Combinatoria de la Sociedad Española de Investigación en Educación Matemática.
- Fischhoff, B., y Kadvan, J. (2011). *Risk: A very short introduction*. Oxford: Oxford University Press.
- Garfield, J., y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Konold, C., y Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy reasoning and thinking* (pp. 169-199). Netherlands: Kluwer Academic Publishers.
- Moore, D. (1990). Uncertainty. En L.A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

- Sánchez, E., Borim, C., y Coutinho, C. (2011). Teachers understanding of variation. En C. Batanero, G. Burril, y C. Reading (Eds.), *Teaching statistics in school mathematics challenges for teaching and teacher education. A Joint ICMI/IASE Study: The 18th ICMI Study* (pp. 211-222). The Netherlands: Springer.
- SEP (2011). *Programas de Estudio. Educación Básica. Secundaria. Matemáticas*. México, D. F.: Secretaría de Educación Pública.
- Shaughnessy, J.M. (2007). Research on statistics learning and reasoning. En F.K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1009). Charlotte, NC: Information Age Publishing.
- Tversky, A., y Kahneman, D. (1991). Loss aversion in riskless choice. A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039-1061.
- Watson, J., Kelly, B., Callingham, R., y Shaughnessy, M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34(1), 1-29.
- Wild, D.J., y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67, 223-265.

Anexo

CUESTIONARIO

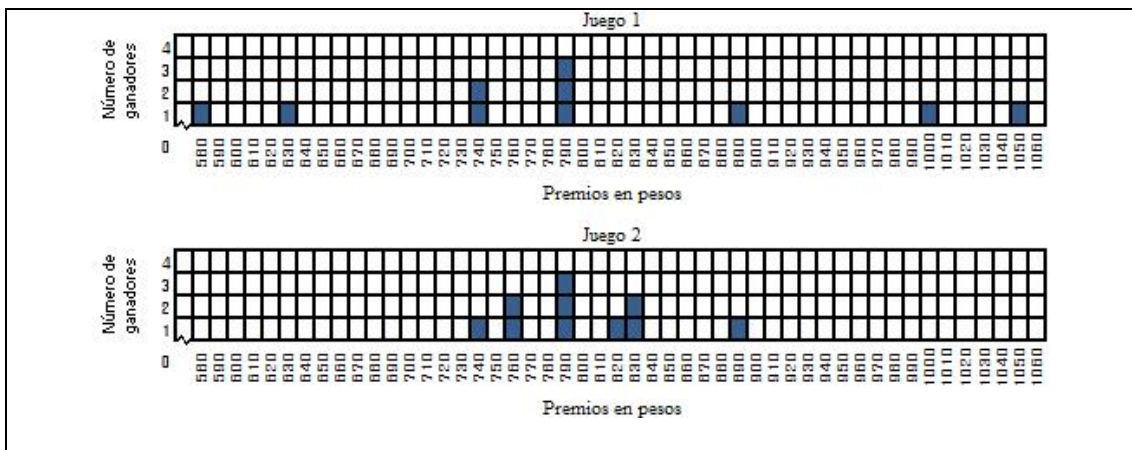
1. En una feria, se invita a los asistentes a participar en dos juegos de apuestas. Considera que participas en uno de esos juegos y puedes elegir en cuál. Los premios en efectivo, que han obtenido diferentes personas se muestran en las Tablas siguientes:

Premios en pesos (Juego 1)	Premios en pesos (Juego 2)
870	1110
945	1140
1110	1140
1110	1185
1185	1185
1185	1185
1185	1230
1335	1245
1500	1245
1575	1335

Contesta lo siguiente

- a) ¿En cuál de los dos juegos participarías? ¿Por qué?
- b) ¿En cuál de los juegos hay más variación? Explica tu respuesta

2. Imagina que participas en uno de dos juegos de apuestas. Antes de aceptar, te permiten conocer los montos de los premios en dinero que puedes ganar. Los premios que han obtenido diferentes personas se muestran en gráficas y son las que se presentan a continuación:



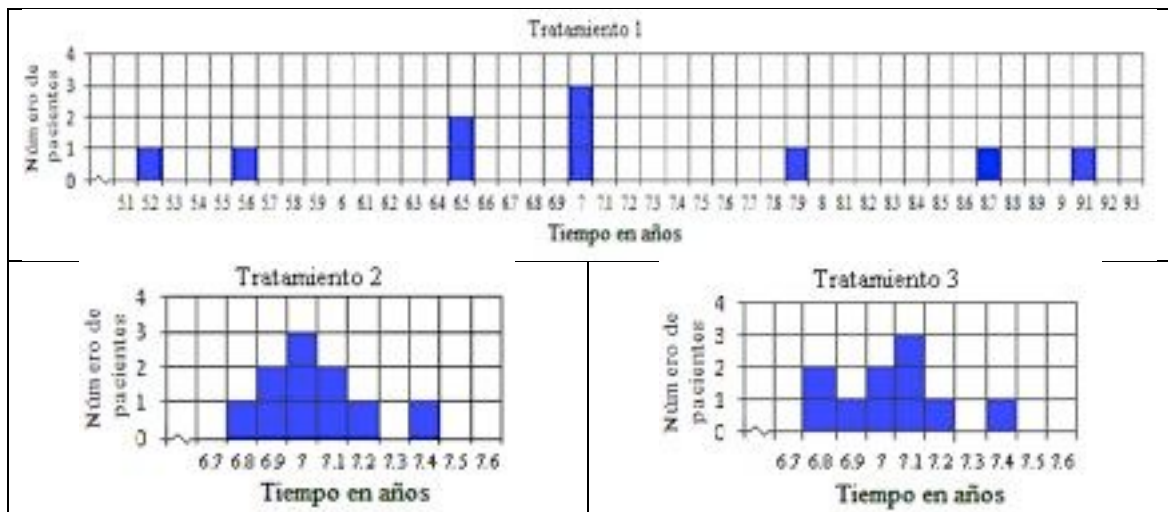
Contesta lo siguiente

- a) ¿En cuál de los dos juegos participarías? ¿Por qué?
- b) ¿En cuál de los juegos hay más variación? Explica tu respuesta

3. Considera que debes aconsejar a una persona que padece una enfermedad grave, la cual es tratable con medicamentos que pueden extender la vida por varios años. Es posible elegir entre tres opciones, dependiendo del tratamiento. Según el tipo de medicina, las personas tienen diferentes reacciones a las sustancias, para algunas, éstas tienen el mismo resultado, mientras que para otras, puede ser mayor o menor. En las Tablas siguientes se muestran los años que han vivido varios pacientes que se han tratado con una de

las opciones mencionadas; cada dato de la Tabla corresponde a un paciente. También, se incluye la representación gráfica de los distintos tratamientos.

Tiempo en años (Tratamiento 1)	Tiempo en años (Tratamiento 2)	Tiempo en años (Tratamiento 3)
5.2	6.8	6.8
5.6	6.9	6.8
6.5	6.9	6.9
6.5	7.0	7.0
7.0	7.0	7.0
7.0	7.0	7.1
7.0	7.1	7.1
7.8	7.1	7.1
8.7	7.2	7.2
9.1	7.4	7.4



Contesta lo siguiente:

- a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? ¿Por qué?
- b) ¿En cuál de los tratamientos hay más variación? Explica tu respuesta

^{xxxii} Esta investigación ha sido subvencionada por: Consejo Nacional de Ciencia y Tecnología (CONACYT), México. Proyecto 101708.