

# INFLUENCIA DEL NIVEL ESCOLAR Y EL CONTEXTO EN EL CONOCIMIENTO INFORMAL DE CONCEPTOS INFERENCIALES

---

ANTONIO MORENO VERDEJO

*I.E.S. "Trevenque", Granada*

ANGUSTIAS VALLECILLOS JIMÉNEZ

*Universidad de Granada*

QUINTO SIMPOSIO DE LA SOCIEDAD ESPAÑOLA DE  
INVESTIGACIÓN EN EDUCACIÓN MATEMÁTICA  
Almería, Septiembre 2001



# INFLUENCIA DEL NIVEL ESCOLAR Y EL CONTEXTO EN EL CONOCIMIENTO INFORMAL DE CONCEPTOS INFERENCIALES



ANTONIO MORENO VERDEJO

*I.E.S. "Trevenque", Granada*

ANGUSTIAS VALLECILLOS JIMÉNEZ

*Universidad de Granada*

## RESUMEN

En este trabajo presentamos parte de los resultados obtenidos en un estudio exploratorio sobre el conocimiento de conceptos básicos en inferencia estadística en estudiantes del nivel de secundaria. La toma de datos se ha llevado a cabo en dos cursos, 3º de E.S.O. y C.O.U., con 49 estudiantes de distinta edad y formación estadística previa. Para ello los estudiantes han contestado un cuestionario escrito que consta de dos partes y 12 preguntas sobre contenidos inferenciales básicos. Los enunciados se plantean en tres contextos distintos, concreto, narrativo y numérico. Sobre las respuestas de los alumnos se han llevado a cabo análisis, fundamentalmente de tipo cualitativo. Por tratarse de un estudio exploratorio nos preocupamos especialmente por determinar las cuestiones abiertas y formular interrogantes a resolver por la investigación experimental posterior más que por sacar conclusiones apresuradas. No obstante, hemos obtenido algunas conclusiones provisionales interesantes.

## ABSTRACT

In this paper we in summary present part of our current results from an exploratory study regarding Spanish secondary school students' knowledge about basic statistical inference concepts. The data were taken from a total of 49 students from two different secondary school level, with different age and previous statistical formation. The student were asked to complete a written questionnaire about basic inferential concepts, which included tasks in three different contexts: concrete, narrative and numerical context. We mainly used qualitative analysis of students' answers. Since this is an exploratory study we mainly intend to determine open questions and to formulate new research problems for further research, instead of trying to reach hasty conclusions. In spite of this we have obtained some preliminary interesting results.

## INTRODUCCIÓN

En este trabajo presentamos parte de los resultados obtenidos en un estudio exploratorio sobre el conocimiento de conceptos básicos en inferencia estadística en estudiantes del nivel de secundaria. Este

estudio exploratorio forma parte del proyecto de investigación “Dificultades teóricas, metodológicas y curriculares de la estadística inferencial en la enseñanza secundaria” que se desarrolla en la actualidad y cuyos propósitos son, en general, determinar áreas problemáticas, dificultades de los alumnos, entornos de aprendizaje favorables o cuestiones metodológicas que puedan servir como base para mejorar la enseñanza de la inferencia estadística en el nivel de secundaria. Algunos otros trabajos son Moreno y Vallecillos (1998; 1999a; 1999b; 2000; en revisión), Vallecillos (1996; 1999), Vallecillos y Moreno (1997).

En nuestro estudio hay algunos presupuestos previos: a) los conceptos estudiados se seleccionan en función de su inclusión en los currículos de secundaria; b) la inferencia estadística es una práctica cultural y c) la comprensión correcta de los conceptos no se produce espontáneamente.

Los actuales currículos de secundaria incluyen el estudio de situaciones de incertidumbre, el estudio de muestras, la representatividad muestral, los procesos de muestreo y las conclusiones que se pueden extraer del estudio de muestras de poblaciones determinadas. Un revisión crítica de este tema puede verse en Moreno (2000).

Las prácticas culturales, entre ellas la inferencia estadística, guían la actividad y facilitan el acceso al conocimiento matemático, indican cómo usar las herramientas. En este sentido, estudiaremos la diferencia en la comprensión y uso de los conceptos seleccionados en diferentes contextos y entre estudiantes con y sin formación estadística.

En Moreno (2000) se recogen diversas investigaciones que afirman que los estudiantes poseen concepciones erróneas sobre la influencia del tamaño de la muestra, sobre la aleatoriedad, sobre la variabilidad y representatividad de una muestra, y sobre la importancia atribuida al tipo de muestreo. Estas concepciones se encuentran muy arraigadas y son difíciles de cambiar.

Las razones anteriores justifican la necesidad de que la investigación educativa se ocupe de analizar el modo en que los alumnos realizan inferencias estadísticas recurriendo al estudio de la comprensión de diversos conceptos implicados y al modo en que los alumnos se enfrentan al propio proceso de inferencia. En nuestro trabajo distinguimos cuatro fases: a) análisis del conocimiento previo de conceptos de inferencia estadística de los estudiantes; b) construcción de un cuestionario específico para la toma de datos; c) administración del cuestionario a estudiantes de dos cursos de nivel de secundaria con diferentes conocimientos estadísticos previos y d) análisis de los resultados con un objetivo exploratorio, determinación de obstáculos, búsqueda de áreas problemáticas, dificultades de los alumnos, etc.

En este trabajo, por razones de espacio, presentamos los resultados que se refieren a los conceptos básicos de población y muestra y a las concepciones sobre el proceso de inferencia de los datos muestrales a la población de referencia.

## METODOLOGÍA

Describimos, resumidamente, los objetivos, la muestra empleada así como el cuestionario y su forma de aplicación.

### *Objetivos*

Los objetivos generales principales de este estudio son: a) evaluar las ideas que poseen los estudiantes del nivel de secundaria sobre conceptos básicos de inferencia estadística; b) estudiar comparativamente estas ideas en alumnos estadísticamente novatos con los que finalizan su formación secundaria y c) analizar la influencia en la conceptualización de distintos contextos de presentación de la pregunta.

Estos objetivos se concretan en los siguientes más específicos:

- Verificar si identifican importantes conceptos inferenciales básicos como la población y muestra de estudio en diferentes contextos;
- Estudiar las concepciones de los alumnos sobre algunas características esenciales de las muestras como la aleatoriedad y variabilidad;
- Explorar las concepciones iniciales de los estudiantes sobre el proceso de muestreo y su validez;
- Estudiar la influencia y el uso que hacen los estudiantes de sus conocimientos previos sobre estadística y probabilidad en los distintos contextos;
- Estudiar si los alumnos reconocen algunos sesgos en el proceso de muestreo y
- Estudiar la influencia que tiene en el proceso de inferencia el tamaño de la muestra estudiada.

### *Sujetos de estudio*

El estudio se ha llevado a cabo con un total de 49 alumnos de dos Institutos de Enseñanza Secundaria de Granada. 30 de ellos pertenecen a 3<sup>er</sup> curso de la E.S.O. y no han recibido formación estadística previa, los otros 19 estudiantes pertenecen al curso de COU, son alumnos que terminan su formación secundaria este curso y pasarán a la Universidad el año próximo. Han recibido alguna formación previa en estadística descriptiva.

### *Cuestionario*

Consta de doce preguntas referidas a los siguientes conceptos de inferencia estadística básica: conceptos de población y de muestra; el proceso de inferencia de la muestra a la población; la influencia del tamaño de la muestra en el proceso de inferencia; la influencia de distintos métodos de muestreo para realizar la inferencia; el concepto de aleatoriedad y la presencia de sesgos en el muestreo. Los enunciados se presentan en tres contextos distintos: a) concreto; b) de narración y c) numérico y está dividido en dos partes: la Parte I que contiene los enunciados en contexto concreto, requiere el empleo de material manipulativo (bolas, botellas de muestreo y baraja de naipes) y la intervención del investigador; la Parte II contiene los enunciados en los contextos narrativo y numérico y no es necesaria la intervención del investigador.

Por razones de espacio en esta comunicación presentamos los resultados que se refieren a los conceptos de población y muestra y al propio proceso de inferencia. Los ítems a los que corresponden estos resultados son los siguientes:

Ítem I.1: Conceptos de población y muestra en contexto concreto

*Tenemos un saco de 100 bolas de dos colores, rojo y verde. Queremos estudiar el número de bolas de cada color. Para ello sacamos 25 bolas y observamos que 14 son rojas y 11 son verdes.*

*Indica:*

*El conjunto de objetos sobre el que se realiza el estudio:*

*La muestra que se observa:*

Ítem II.1: Conceptos de población y muestra en contexto narrativo

*Para conservar alimentos en casa la mayoría de las veces se utilizan tarros de mayonesa para evitar comprar tarros especiales. La revista Conservas quiere saber que porcentaje de tarros de mayonesa se rompen cuando se utilizan para conservas. La revista consiguió 10 tarros de ma-*

*yonesa para conservar tomate. Solamente tres de ellos se rompieron.*

*Identifica:*

*El conjunto de objetos sobre el que se realiza el estudio:*

*La muestra que se observa:*

Ítem II.2: Conceptos de población y muestra en contexto numérico

*El Ayuntamiento se plantea quitar un parque para construir unos aparcamientos. Se desea conocer la opinión de los habitantes mayores de 18 años y para ello se encuesta a 450 de ellos. El 70% no desea que se quite el parque.*

*Indica: La población de la que se desea obtener información:*

*La muestra que se observa:*

Ítem I.2: Proceso de inferencia en contexto concreto

*Tu profesor acaba de enseñarte una muestra formada por cinco cartas. Sabiendo que en el mazo hay 30, ¿cuántas crees que serán rojas?. Indica la razón.*

Ítem II.3: Proceso de inferencia en contexto numérico

*El Ayuntamiento de Granada ha iniciado una campaña explicando que se debe hacer cuando queremos deshacernos de muebles en mal estado. Quiere saber si las instrucciones han resultado claras y comprensibles. La población de Granada es 300.000 personas así que decide preguntar a 2000 adultos de Granada sobre lo que piensan. Han preguntado a personas de barrios grandes y de barrios pequeños, algunos hombres y algunas mujeres, algunos jóvenes y otros más mayores, y a algunos que viven en casas y algunos que viven en pisos. Están seguros de que tienen un variado grupo de personas. El resultado de la encuesta es que el 73% de estos adultos creen que las instrucciones del Ayuntamiento son claras y el 27% piensan que no lo son.*

*¿Qué puedes decirle al Ayuntamiento sobre el porcentaje de adultos en toda Granada que piensa que las instrucciones son claras?*

*a. El 50% porque probablemente la mitad de las personas pensaron que las instrucciones eran claras y la mitad de ellos pensó que no lo eran.*

*b. El 73% porque los adultos a los que se les preguntó dan una idea general de lo que ocurriría si se preguntase a toda la población.*

*c. No puedo decir nada porque el resultado de la encuesta podría haber sido cualquier otro.*

*d. No puedo decir nada porque no pueden preguntar a todos los adultos de la ciudad.*

*e. Sería el \_\_\_\_\_ porque \_\_\_\_\_*

## *Procedimiento de aplicación*

Los estudiantes de cada grupo completaron el cuestionario al mismo tiempo siguiendo las indicaciones del investigador. El grupo de COU contestó el cuestionario en una sesión de 60 minutos, mientras el grupo de 3º de ESO necesitó una sesión para la Parte I del cuestionario y una segunda sesión para la Parte II.

En el ítem I.2 el investigador interviene modo que se describe a continuación: extrae al comienzo de la sesión 22 cartas de un mazo de 52 y, sin mirarlas, son apartadas del resto. Se explica entonces a los alumnos que nos proponemos conocer la proporción de cartas de rojas y negras que contiene el mazo de

30 cartas que nos quedan y para ello se realiza una extracción de cinco cartas, se enseña la muestra a los alumnos que apuntan su composición y se les pide que contesten a continuación el ítem 2.

Durante la sesión, los alumnos trabajaron el cuestionario de forma individual. Se evitó introducir en el cuestionario términos técnicos (población, muestreo, etc.) pero si los alumnos tenían alguna duda se resolvía de forma individual evitando que la explicación orientara la respuesta. Si aún así, el estudiante continuaba sin comprender la cuestión, se le pedía que la dejase sin contestar.

## ANÁLISIS DE LOS RESULTADOS

Por razones de espacio recogemos aquí los resultados que se refieren sólo a los conceptos de población y muestra y al propio proceso de muestreo.

### *Conceptos de población y muestra*

En la Tabla 1 tenemos resumidos los datos acerca de la identificación de la población y la muestra.

Identifican la población de estudio en el contexto concreto el 57.1% de todos los estudiantes, mientras que sólo lo hacen el 49% en los contextos narrativo y el numérico.

Identifican la muestra el 30.6% de los estudiantes en el contexto concreto, el 22.4% en el contexto narrativo y el 20.4% en el numérico.

No identifican la población ni la muestra en el contexto concreto el 42.9%, el 48.9% en el narrativo y el 49% en el contexto numérico.

En estos primeros resultados globales se observa un mayor porcentaje de alumnos que identifican la población o la muestra o no identifican ninguna de ellas en el contexto concreto que en el narrativo y el numérico. Parece deducirse de ellos claramente una ordenación de los contextos, concreto, narrativo, numérico, de mayor a menor facilidad para los alumnos en cuanto a la identificación de los conceptos analizados se refiere.

En el estudio por cursos, el 53.3% de los alumnos de ESO identifican la población frente al 63.2% de los de COU en el contexto concreto; el 50% de los alumnos de ESO lo hacen frente al 47.4 % de los de COU en el contexto narrativo y en el contexto numérico los resultados obtenidos son el 33.3% en el grupo de ESO y el 73.6% de los alumnos de COU.

Los resultados para el grupo de ESO (53.3%, 50.0%, 33.3%) mantienen el orden inicial para los contextos (concreto, narrativo, numérico) de mayor a menor facilidad. Para el grupo de COU (63.2%, 47.4%, 73.6%), sin embargo se deduce que ha resultado para ellos más familiar y más fácil el contexto numérico para la identificación de la población estudiada. Comparativamente, los porcentajes de alumnos que identifican la población, son mas altos para los alumnos de COU en los contextos concreto y numérico, con una diferencia especialmente notable para el caso numérico. En el narrativo son ligeramente más altos los porcentajes del grupo de ESO (50%, 47.4%).

Tabla 1: Identificación de población y muestra

Frecuencia % Grupo	ESO			COU		
	Id. Población	Id. Muestra.	NP / NM	Id. Población	Id. Muestra	NP / NM
Concreto	14 53.3	7 23.3	14 46.7	12 63.2	8 42.1	7 36.8
Narrativo	15 50.0	5 16.7	15 50.0	9 47.4	6 31.6	9 47.3
Numérico	10 33.3	3 10.0	20 66.6	14 73.6	8 42.1	4 21.1

Si analizamos la identificación de la muestra por cursos observamos los datos siguientes: en el contexto concreto lo hacen el 23.3% de los alumnos de ESO frente al 42.1% de los de COU; en el narrativo el 16.7% de alumnos de ESO frente al 31.6% en COU; en el contexto numérico lo hacen el 10% en ESO y el 42.1% en COU.

Los resultados para el grupo de ESO (23.3%, 16.7%, 10%) mantienen el orden citado (concreto, narrativo, numérico) para los contextos de mayor a menor facilidad. En COU los resultados (42.1%, 31.6%, 42.1%) sin embargo indican pocas diferencias entre los contextos. Comparativamente, los resultados obtenidos por el grupo de ESO son considerablemente más bajos que los de COU. También las diferencias, con respecto al caso de la identificación de la población, son considerablemente más grandes que las anteriores. Estos datos son un indicador inicial de una posible mayor dificultad para la identificación de la muestra que la de la población en todos los contextos y por cursos.

No identifican la población ni la muestra, por cursos, el 46.7% de los alumnos de ESO frente al 36.8% de los de COU en el contexto concreto; el 50% en ESO frente al 47.3% en COU en el narrativo y el 66.6% en ESO frente al 21.1% en COU.

Comparando nuevamente por contextos, los resultados para ESO son (46.7%, 50%, 66.6%) y para COU (36.8%, 47.3%, 21.1%) observamos que las diferencias son ciertamente notables, excepto para el contexto narrativo. La diferenciación mayor se produce en el contexto numérico, en donde se observa una diferencia entre los porcentajes del 45%.

En resumen: muchos estudiantes de ambos niveles de enseñanza no identifican los conceptos básicos de población y muestra si bien hay diferencias importantes en los porcentajes de éxito en los ítems correspondientes en función del contexto. El concreto es el que ofrece mayor porcentaje de respuestas correctas. El grupo de COU ofrece mejores resultados globales que el de ESO y en el contexto numérico. Las dos terceras partes de los alumnos de ESO no identifican ni la población ni la muestra mientras que en COU no lo hacen la quinta parte.

### *Proceso de inferencia de la muestra a la población*

Para categorizar las respuestas de los alumnos se utilizó inicialmente la clasificación en tres grandes grupos según los *criterios de descripción de la población* empleados en ellas descrita en Moreno (2000). Estos fueron:

C1) *Concepción inferencial*: El proceso de inferencia está sujeto al azar y no permite determinar con precisión las características de una población a partir de las de una de sus muestras.

C2) *Concepción de identidad*: El proceso de inferencia permite describir la población con características idénticas a las de una de sus muestras.

C3) *Concepción previa*: La población tiene unas características descritas por ideas previas y no por las observadas en la muestra extraída.

En el contexto narrativo hemos encontrado hasta un 20.4% del total de respuestas que expresan la idea de que la encuesta no permite saber nada si no se estudia toda la población. Hemos añadido, pues, una nueva categoría en la clasificación anterior que hemos llamado *concepción determinista*:

C4) *Concepción determinista*: Solo el estudio de la población completa permite determinar sus características propias.

En la Tabla 2 están resumidos los resultados obtenidos según los contextos y por grupos de estudiantes. Como puede verse en ella, en los dos contextos y en los dos grupos la mayoría de los alumnos manifiestan la *concepción de identidad*, C2, en el proceso de inferencia. Se manifiesta independiente del

contexto y en porcentajes altos, en general, y especialmente en COU. En este curso el porcentaje de alumnos que manifiestan esta concepción es especialmente alto en el contexto concreto.

El número de respuestas en blanco es mayor en el contexto concreto, 18.4%, que en el contexto narrativo, 10.2%.

Por cursos, los porcentajes de aparición de C2 en ambos contextos están en torno al 60% en ESO y entre el 60 y el 87% en COU.

La concepción C4 aparece en el 23.1% de las respuestas de los alumnos de ESO y en un 22.2% de las de COU.

En resumen: muchos alumnos manifiestan la *concepción de identidad* en sus respuestas acerca del proceso de inferencia que es una manifestación explícita de la heurística de la representatividad descrita por Kahneman y cols. (1982) en otro contexto.

Tabla 2: Concepciones sobre el proceso de inferencia

Frecuencia % Grupo	ESO				COU			
	C1	C2	C3	C4	C1	C2	C3	C4
Concreto		15 60.0	10 40.0			13 86.7	2 13.3	
Narrativo	5 19.2	15 57.7		6 23.1	3 16.3	11 61.1		4 22.2

## CONCLUSIONES

En este estudio hemos llevado a cabo una exploración organizada de los conocimientos que poseen dos grupos de estudiantes de nivel de secundaria sobre unos conceptos inferenciales básicos para el aprendizaje formal de la inferencia estadística en este nivel de enseñanza.

Hemos encontrado importantes errores que afectan a los conceptos de población y muestra en todos los contextos en los que se han presentado las preguntas correspondientes. Estos resultados son llamativos porque podría suponerse que, al tratarse de conceptos subyacentes en muchas actividades de la vida diaria en diversas formas, los estudiantes tienen muchos ejemplos y situaciones que les sirven de base para construir estos conceptos.

El estudio de una muestra no proporciona una buena estimación de las características de la población estudiada, en opinión de muchos alumnos. Tampoco parecen apreciar claramente las ventajas de la aleatoriedad frente a otros tipos de muestreo.

Todos estos resultados han sido obtenidos de un estudio exploratorio sobre una muestra intencional de estudiantes y no pueden ser generalizados en absoluto. Sin embargo, dibujan un panorama de concepciones e ideas previas de los estudiantes de secundaria que es necesario tener en cuenta por las implicaciones de todo tipo que de ellas se derivan. En nuestro caso nos han servido de momento para generar una serie de hipótesis de investigación sobre las que es necesario trabajar mucho todavía con el fin de diseñar una metodología de trabajo para la enseñanza basada en teorías generadas por la observación sistemática y organizada de la enseñanza y el aprendizaje de la estadística inferencial en este nivel de enseñanza.

## AGRADECIMIENTOS

Este trabajo se realiza en el marco del Proyecto de Investigación PB97-0827 de la Dirección General de Enseñanza Superior, MEC, Madrid.

## REFERENCIAS

- Kahneman, D., Slovic, P., y Tversky, A. (eds.). (1982). *Judgment under uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Moreno, A. (2000). *Investigación y enseñanza de la estadística inferencial en el nivel de secundaria*. Memoria de Tercer Ciclo. Universidad de Granada.
- Moreno, A. y Vallecillos, A. (1998). El muestreo en la enseñanza secundaria. En F. Muñoz, D. Cárdenas y A. López (Eds.): *Actas de las VIII Jornadas Andaluzas de Educación Matemática "Thales"*, (pp. 249-254). Jaén: S.A.E.M. "Thales".
- Moreno, A. y Vallecillos, A. (1999a). ¿Cuántas ranas hay en la charca?. *Actas de las 9º Jornadas para el Aprendizaje y la Enseñanza de las Matemáticas*, (pp. 294-296). Lugo: CEFOCOP.
- Moreno, A. y Vallecillos, A. (1999b). La educación estadística en la sociedad actual. En: I. Berenguer, J. M. Cardeñoso y M. Toquero (Eds.): *Investigación en el aula de matemáticas: Matemáticas en la sociedad*, (pp 253-261). Granada: Universidad de Granada y Sociedad Andaluza de Educación Matemática "Thales".
- Moreno, A. y Vallecillos, A. (en revisión). Exploración de concepciones iniciales sobre procesos inferenciales. *Educación Matemática*.
- Moreno, A.; Vallecillos, A. (2000). Dificultades en la comprensión de conceptos básicos de inferencia estadística en la educación secundaria. En: A. Gámez, C. Macías y C. Suárez (Eds.): *Actas del IX Congreso Andaluz de Educación Matemática "Thales"*, (pp. 109-111). Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Vallecillos, A. (1996). *Inferencia Estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Granada: Comares.
- Vallecillos, A. (1999). Some empirical evidences on learning difficulties about testing hypotheses. Ponencia invitada, Topic IPM 58. *Proceeding of the 52<sup>nd</sup> Session of the International Statistical Institute*, Vol. 2, Tome LVIII, pp. 201-204. The Netherland: ISI.
- Vallecillos, A. y Moreno, A. (1997). Los profesores de matemáticas y la inferencia estadística en la enseñanza secundaria. En I. Berenguer, B. Cobo y F. Fernández (Eds.): *Investigación en el aula de Matemáticas: La tarea docente*, (pp. 279-287). Granada: Universidad de Granada y Sociedad Andaluza de Educación Matemática "Thales".