

Reflexiones sobre el uso de la estadística inferencial en investigación didáctica
Reflections on the use of inferential statistics in data analysis during a didactic
research

Reflexões sobre o uso de estatísticas inferenciais na análise de dados durante
uma investigação didática

Nicolás Veiga. ORCID: 0000-0001-8552-058X¹

Lucía Otero. ORCID: 0000-0002-3569-1660¹

Julia Torres. ORCID: 0000-0001-8257-6334¹

¹Universidad de la República, Uruguay. Contacto: nveiga@fq.edu.uy

Resumen:

Como producto de procesos investigativos en educación se generan gran cantidad de datos cuyo análisis representa un reto para los investigadores. La enorme variedad de factores interrelacionados que pueden tener incidencia en los datos dificulta sustancialmente la extracción de conclusiones confiables, tanto a la hora de describir o predecir el comportamiento de una población estudiantil como a la de evidenciar el efecto de intervenciones didácticas. El análisis estadístico inferencial provee herramientas que permiten la evaluación sistemática y eficiente de una muestra de la población que se quiere estudiar. Presentamos aquí una reflexión sobre la operativa concreta e interpretación de resultados asociados a la aplicación de tres instrumentos inferenciales (distribución de probabilidad, intervalo de confianza y test de hipótesis) a un caso de estudio derivado de nuestra experiencia docente. Estas herramientas nos permitieron evaluar la incidencia de la modalidad de cursado (presencial o en línea) en el desempeño académico de estudiantes de primer año de la Facultad de Química, Universidad de la República. Además, estos mismos instrumentos estadísticos se aplicaron al estudio de la influencia en la misma población de una intervención didáctica, en la que se incorporaron a un curso masivo de química de primer año (Química General I) un conjunto de materiales con el objetivo de mejorar el desempeño académico mediante el aprendizaje activo y la autorregulación. En este sentido, la estadística inferencial hizo posible el análisis sistemático y eficiente de la información derivada en este estudio de caso, lo que permitió arribar a conclusiones que son estadísticamente significativas.

Palabras clave: análisis inferencial, herramientas estadísticas, análisis de datos, investigación didáctica.

Abstract:

As a product of educational research activities, a large amount of data is produced, the analysis of which represents a challenge for the researchers. In this context, the large number of interrelated factors that may have an impact on the data makes it extremely difficult to draw reliable conclusions, either when describing or predicting the behavior of a student population, or when evidencing the effect of didactic interventions. Statistical inferential analysis provides tools that allow the systematic and efficient evaluation of a sample of the population to be studied. In this work we present a reflection on the implementation and interpretation of results associated with the application of three inferential instruments (probability distribution, confidence interval and hypothesis test) to a case study derived from our teaching experience. These instruments allowed us to evaluate the influence that the course modality (face-to-face or online) exerts on the academic performance of first-year students of Facultad de Química, Universidad de la República. In addition, these statistical tools were also applied to the study of the incidence, on the same population, of a didactic intervention in which a set of materials was incorporated into a massive

first-year chemistry course (General Chemistry I), with the aim of promoting active learning and self-regulation. In this regard, inferential statistics allowed the systematic and efficient analysis of the information associated with this case study, allowing to draw statistically significant conclusions.

Keywords: inferential analysis, statistical tools, data analysis, didactic research.

Resumo:

Como produto dos processos de pesquisa educacional, uma grande quantidade de dados é gerada, cuja análise representa um desafio para os pesquisadores. Nessa estrutura, o grande número de fatores inter-relacionados que podem ter impacto nos dados torna extremamente difícil tirar conclusões confiáveis, tanto ao descrever ou prever o comportamento de uma população estudantil, por exemplo, quanto ao evidenciar o efeito de intervenções didáticas. A análise estatística inferencial fornece ferramentas que permitem a avaliação sistemática e eficiente de uma amostra da população a ser estudada. Neste trabalho, apresentamos uma reflexão sobre a operação e interpretação concretas dos resultados associados à aplicação de três instrumentos inferenciais (distribuição de probabilidade, intervalo de confiança e teste de hipóteses) a um estudo de caso derivado de nossa experiência de ensino. Essas ferramentas de análise nos permitiram avaliar a incidência da modalidade do curso (presencial ou online) no desempenho acadêmico dos alunos do primeiro ano da Faculdade de Química da Universidad de la República. Além disso, esses mesmos instrumentos estatísticos foram aplicados ao estudo da influência sobre a mesma população de uma intervenção didática, na qual um conjunto de materiais foi incorporado a um curso maciço de química do primeiro ano (Química Geral I), com o objetivo de melhorar o desempenho acadêmico através da aprendizagem ativa e auto-regulação. Nesse sentido, a estatística inferencial possibilitou a análise sistemática e eficiente das informações derivadas neste estudo de caso, permitindo chegar a conclusões estatisticamente significativas.

Palavras-chave: análise inferencial, ferramentas estatísticas, análise de dados, pesquisa didática.

Fecha de recibido: 27/07/2020

Fecha de aceptado: 03/10/2020

Introducción

En el desarrollo de cualquier investigación en educación es frecuente recabar múltiples datos para la identificación y diagnóstico de situaciones y problemas particulares en el contexto de estudio, así como para la evaluación de los resultados de las intervenciones. Estos datos, obtenidos mediante la observación y/o la realización de distintas evaluaciones, encuestas, sondeos, etcétera, son muchas veces difíciles de analizar debido a su gran número y a la complejidad del entramado multicausal que modula la información que se quiere extraer (Gamboa, 2018; Maguire y Rowley, 1992).

Para analizar de forma eficiente y sistemática una gran cantidad de datos es necesario utilizar en forma adecuada las distintas herramientas que nos provee la estadística. Por un lado, las herramientas descriptivas como estimadores, tablas de contingencia, histogramas y diagramas de caja permiten resumir y comunicar el comportamiento de una muestra en particular de forma holística y narrativa, identificando patrones y tendencias que solo son aplicables a la muestra estudiada (Veiga, Otero y Torres, 2018). Sin embargo, cuando se requiere inferir un comportamiento general a partir de los datos obtenidos para una muestra en particular, es necesario emplear las herramientas de la estadística inferencial (Agresti, 2018).

En este trabajo se presenta la aplicación de varios instrumentos inferenciales (distribución de probabilidad, intervalo de confianza y test de hipótesis) a un caso de estudio derivado de una investigación didáctica realizada en la Facultad de Química, Universidad de la República (Udelar),

y se aportan detalles sobre su aplicación concreta, así como un análisis reflexivo sobre las fortalezas y limitaciones de este tipo de análisis.

Caso de estudio

Los estudiantes de primer año de la Facultad de Química de la Udelar cursan, en el primer semestre de todas las carreras, la asignatura Química General I. Esta asignatura puede cursarse en dos modalidades: presencial o en línea. Como primer paso en la aplicación concreta de las herramientas estadísticas inferenciales al caso de estudio, es necesario definir la población y la muestra que se va a explorar. La población elegida, es decir el conjunto total de sujetos que queremos estudiar, son “los estudiantes que cursan primer año en la Facultad de Química de la Udelar”. Esta población resulta adecuada a nuestros objetivos, ya que pretendemos inferir el desempeño académico estudiantil de forma atemporal y genérica para poder tomar decisiones de forma independiente de la coyuntura social, económica, cultural, etcétera, de la o las generaciones que podamos tomar como muestra. De esta manera, la población de estudiantes a analizar es muy extensa (tendiendo a infinito), lo que trae como consecuencia que, en la práctica, no la podamos estudiar de forma completa. Una solución es analizar una muestra o subconjunto de la población a estudiar, tal como podría ser una o varias generaciones de todos esos estudiantes (Gamboa, 2018). Para nuestro caso de estudio, se seleccionaron como muestras, por un lado, la generación 2013 y, por otro, el conjunto completo de estudiantes de las generaciones 2012 a 2018.

Elegiremos además para el caso de estudio una característica o variable a medir en las muestras para luego inferir qué valor podría tener esa característica en la población y realizar así predicciones. En este caso, emplearemos como variable el desempeño académico. Esta variable será cuantificada mediante el *puntaje* (sobre 40 puntos) que obtienen los estudiantes en la primera evaluación sumativa del curso, la cual presenta formato múltiple opción y posee un nivel equivalente año a año, ya que se genera a partir de una base de datos que contiene ejercicios de dificultad similar.

El primer objetivo del análisis inferencial realizado fue evaluar el desempeño académico de la población y la incidencia en él de la modalidad de cursado (presencial o en línea). En segundo lugar, el análisis fue empleado para inferir el comportamiento de la población descrita en cuanto a su desempeño académico, en el marco de una intervención didáctica realizada sobre una muestra de la población con el objetivo de disminuir la deserción mediante la mejora del desempeño académico en los estudiantes de primer año. Si bien la deserción universitaria es multicausal (Arim y Katzkowicz, 2017), la falta de motivación extrínseca en los casos de bajos desempeños académicos y la no existencia de una capacidad autorregulatoria adecuadamente desarrollada se señalan como factores a atender (Gravina y Prieto, 2019). En este contexto, en los años 2017 y 2018 se realizó una intervención en el curso de Química General I tendiente a favorecer el aprendizaje activo y la autorregulación con el fin de mejorar el desempeño académico de los estudiantes. Para ello se trabajó en la implementación de nuevos materiales de estudio interactivos (Imer et al., 2018) para ser usados por la totalidad de los estudiantes (tanto los que cursan en modalidad presencial como aquellos que lo hacen en línea). Los nuevos materiales contienen herramientas interactivas activables por el usuario en distintos niveles de dificultad, que ofrecen guías de resolución de ejercicios y devoluciones automáticas instantáneas para las diferentes opciones, entre otras aplicaciones.

Metodología

Los puntajes obtenidos por los estudiantes fueron extraídos de la base de datos del departamento docente y procesados en una planilla de cálculo electrónica de Excel® con el complemento RealStatistics®.ⁱ Este paquete es de distribución gratuita y permite la aplicación de una amplia variedad de herramientas descriptivas e inferenciales a bases de datos extensas de forma intuitiva y rápida (Zaiontz, 2020). Los datos de la generación de estudiantes del año 2013 y su análisis estadístico descriptivo se obtuvieron de un trabajo previo de nuestro grupo (Veiga, Otero y Torres, 2018). Durante el análisis inferencial, los valores atípicos no fueron considerados en el cálculo de los estimadores.

Aplicación de las herramientas estadísticas inferenciales

Distribución de probabilidad: población vs. muestra

El objetivo de la aplicación de la estadística inferencial fue inferir cómo se comporta la variable *desempeño académico* en la población (los estudiantes que cursan primer año en la Facultad de Química de la Udelar) a partir de la información contenida en una muestra de esa población. Para esto es necesario conocer con qué probabilidad ocurren los diversos valores de la variable en la población (*distribución de la población*). En nuestro caso en particular, la población consta de los estudiantes universitarios de primer año considerados de forma atemporal y genérica, por lo que no tenemos acceso a todos ellos. Sin embargo, podemos asumir que la distribución de la población es normal, dependiente de los parámetros μ (valor medio del puntaje) y σ (desviación estándar), y que por tanto el 68 % de los puntajes se encuentran en el intervalo $\mu \pm \sigma$ (desviación estándar) y que al 95 % de ellos los podemos encontrar en el rango $\mu \pm 1,96\sigma$ (Agresti, 2018, pp. 79-92). A los parámetros μ y σ los podemos *estimar* a su vez mediante el promedio (\bar{x}) y la desviación estándar muestral ($s = \sqrt{\sum(x_i - \bar{x})^2 / (n - 1)}$), respectivamente, donde n es el número de elementos de la muestra (Agresti, 2018, pp. 53-58). El grado en que esos estimadores reflejan el comportamiento de la población es uno de los puntos centrales en el análisis inferencial. Este tipo de análisis estadístico permite realizar inferencias sobre la población en base a estos estimadores calculados con los datos de la muestra y determinar qué nivel de confianza tienen estas predicciones.

Volviendo a nuestro caso de estudio, como primer paso en el análisis inferencial estaremos la distribución de probabilidad de la población tomando como muestra en primer lugar a los estudiantes de la generación 2013 ($n = 479$). El histograma de la frecuencia con la que se obtuvo cada puntaje en esta generación se muestra en la figura 1 (Veiga, Otero y Torres, 2018). A partir de estos resultados, se pueden analizar la distribución de los puntajes, su promedio ($\bar{x} = 18,3$) y la desviación estándar muestral ($s = 6,4$). Se puede ver que se observan además *valores atípicos*, que no contribuyen a la distribución general de los datos (destacados en la figura 1).

El histograma de la figura 1 posee una distribución similar a una normal, debido a que el número de sujetos en la muestra ($n = 479$), aunque es grande, no tiende a infinito, por lo que algunos puntajes tienen frecuencias que están apartadas de la tendencia de la curva normal. La tabla 1 muestra el valor de los estimadores \bar{x} y s cuando se excluyen los valores atípicos. A partir de ellos se puede graficar una estimación de la curva normal de la población ($f(x)$), la que se muestra superpuesta al histograma de la muestra en la figura 1. A partir de esta curva normal podemos estimar la probabilidad de que, al elegir aleatoriamente un estudiante de la población, su puntaje esté en un intervalo dado. Por ejemplo, hay un 68 % de probabilidad de que “un estudiante de primer año de Facultad de Química” obtenga un puntaje entre 12,1 ($18,2 - 6,1$) y 24,3 ($18,2 + 6,1$). Por supuesto, la confianza que tenemos en esta afirmación aumenta cuanto más grande es la muestra empleada en la estimación, ya que su distribución de probabilidad se acercará más a la de la población (Agresti, 2018, pp. 79-92).

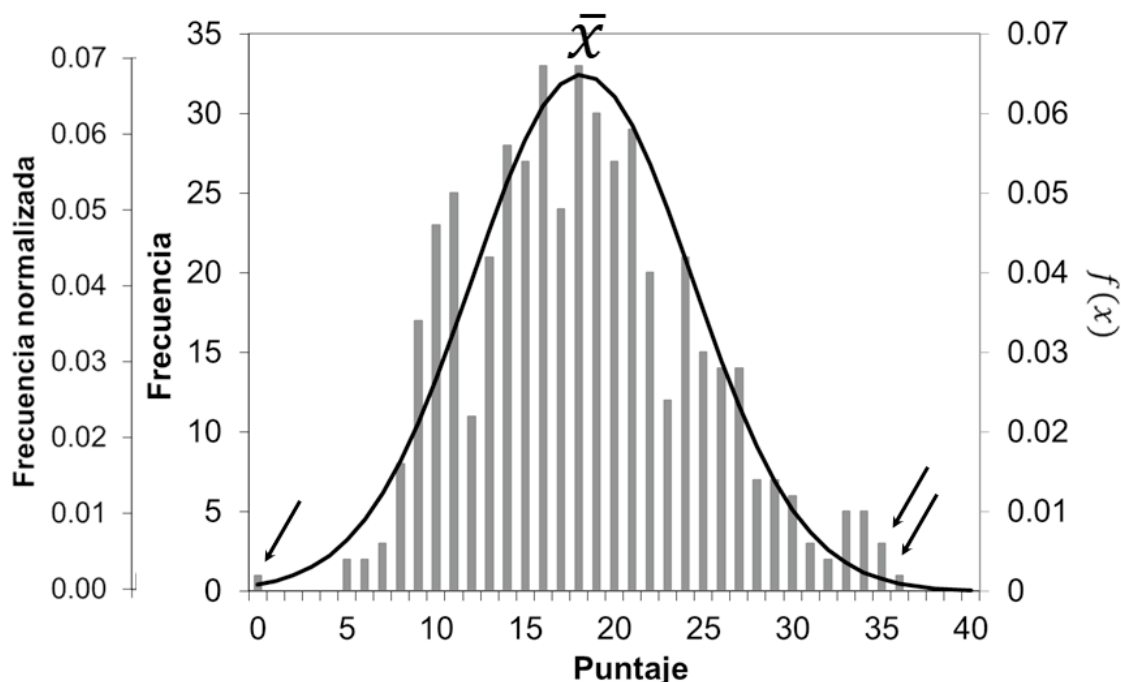


Figura 1: Histograma de la frecuencia con la que se obtuvo cada puntaje para los estudiantes de la generación 2013. La escala de frecuencia normalizada corresponde al cociente entre la frecuencia y el número total de estudiantes. Se muestra superpuesta la distribución normal estimada $f(x)$ de la población. Se señalan asimismo los valores atípicos, \bar{x} = promedio muestral.

	Generación 2013 (Veiga, Otero y Torres, 2018)			Generaciones 2012-2018			Generaciones 2012-2016	Generaciones 2017-2018
	Todos	Presenciales	En línea	Todos	Presenciales	En línea		
\bar{x}	18,2	18,5	16,6	19,1	19,2	17,0	18,9	19,6
μ	$18,2 \pm 0,6$	$18,5 \pm 0,6$	$16,6 \pm 1,5$	$19,1 \pm 0,2$	$19,2 \pm 0,2$	$17,0 \pm 0,8$	$18,9 \pm 0,3$	$19,6 \pm 0,4$
IC	17,6 – 18,8	17,9 – 19,1	15,1 – 18,1	18,9 – 19,3	19,0 – 19,4	16,2 – 17,8	18,6 – 19,2	19,2 – 20,0
s	6,1	6,4	4,5	7,0	7,0	5,9	6,9	7,3
n	474	442	36	3.746	3.513	230	2.661	1.086

Tabla 1: Promedio (\bar{x}), valor medio (μ) y su intervalo de confianza (IC), desviación estándar (s) y número de estudiantes (n) para las generaciones de estudiantes consideradas. Los cálculos se realizaron sobre los puntajes (intervalo = 0-40), sin considerar los valores atípicos.

Intervalos de confianza: descripción y comparación de poblaciones

En base a lo discutido en la sección anterior, solo podemos estimar el puntaje medio de la población, μ , mediante el promedio de la muestra $\bar{x} = 18,2$, pero no podemos conocer cuán buena es esta estimación. No obstante, lo que sí se puede hacer es calcular un intervalo de valores centrado en \bar{x} , que con un cierto nivel de confianza contenga a μ .

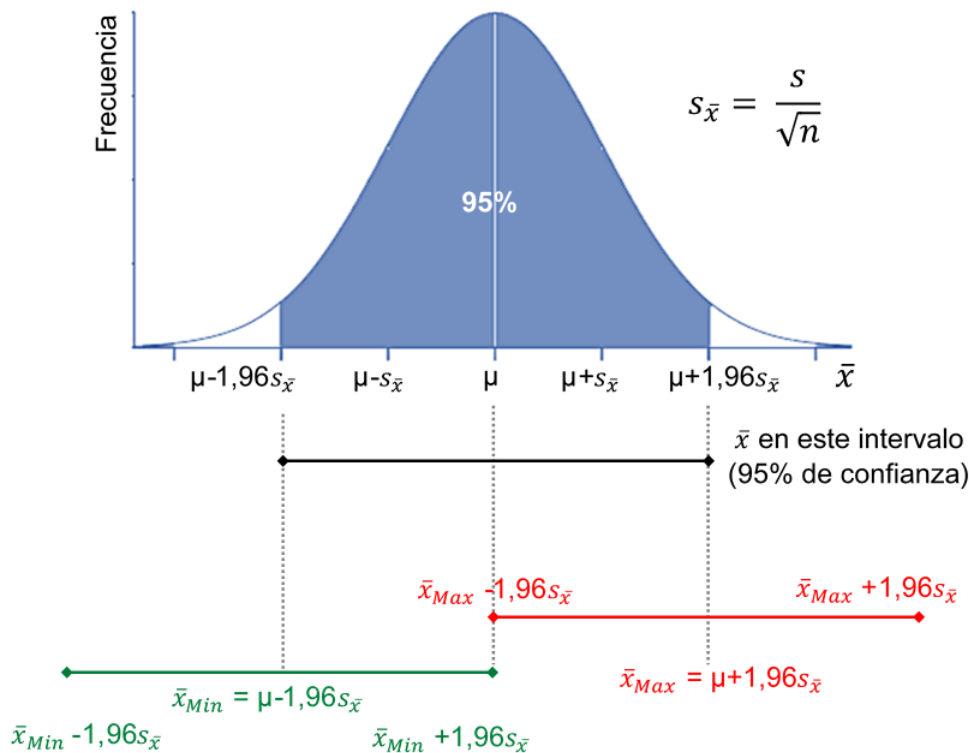


Figura 2: Distribución de probabilidad del promedio muestral \bar{x} (eje $x = \bar{x}$; eje $y =$ frecuencia o probabilidad de ocurrencia; $\mu =$ valor medio poblacional; $s =$ desviación estándar de la muestra; $s_{\bar{x}} =$ error estándar; $n =$ tamaño de la muestra).

Para esto observemos primero la curva de distribución del promedio \bar{x} (figura 2), es decir, la frecuencia (o probabilidad) con la que se obtendrían los valores de \bar{x} al hacer el promedio de un número muy grande (tendiendo a infinito) de muestras distintas de estudiantes, pero todas del mismo tamaño n . Esta distribución de los promedios es normal (como la de la figura 1), centrada en el valor medio de la población (μ), pero con una desviación estándar ($s_{\bar{x}}$; error estándar) menor que la de cualquiera de las muestras: $s_{\bar{x}} = s/\sqrt{n}$. Dado que es una curva normal, el puntaje promedio (\bar{x}) de una muestra cualquiera de n estudiantes de la población puede tomar distintos valores, pero está comprendido, con un 95 % de probabilidad, entre $\mu - 1,96s_{\bar{x}}$ y $\mu + 1,96s_{\bar{x}}$. Podemos concluir entonces que el valor medio de la población, μ , se encuentra (con un 95 % de probabilidad) como mínimo $1,96s_{\bar{x}}$ por debajo el promedio ($\bar{x}_{Max} - 1,96 s_{\bar{x}}$) y como máximo $1,96s_{\bar{x}}$ por encima del promedio ($\bar{x}_{Min} + 1,96 s_{\bar{x}}$). Por lo tanto, μ se encontrará, con un 95 % de confianza, en el rango $\bar{x} \pm 1,96s_{\bar{x}}$, el que se denomina *intervalo de confianza (IC)*. De esta manera, a partir de parámetros de la muestra ($\bar{x}, s_{\bar{x}}$) podemos inferir en qué rango de valores podrá estar el valor medio (μ) de la población. Esta predicción será más precisa a medida que aumentemos el tamaño de la muestra (n), ya que $s_{\bar{x}}$ (y por tanto el *IC*) será menor.

Apliquemos el intervalo de confianza a nuestro caso de estudio. Para la muestra (generación 2013), $\bar{x} = 18,2$ y $s_x = 0,28$ (tabla 1), por lo que, con un 95 % de confianza, μ se encuentra entre 17,6 ($18,2 - 1,96s_{\bar{x}}$) y 18,8 ($18,2 + 1,96s_{\bar{x}}$). De esta forma, inferimos que el estudiante promedio de la población no supera el 50 % del puntaje de la prueba, ya que 20 (mitad del puntaje máximo) queda fuera del *IC* de μ . Ahora bien, si realizamos el mismo procedimiento, pero en lugar de usar una muestra de una sola generación empleamos varias generaciones (por ejemplo la totalidad de los estudiantes que cursaron de 2012 a 2018, intervalo que incluye a la generación 2013, tabla 1), tendremos un mejor estimador del valor medio ($\bar{x} = 19,1$) y con un intervalo de confianza más estrecho (18,9 – 19,3). Recordemos que $s_{\bar{x}} = s/\sqrt{n}$, por lo que, si aumentamos el número de estudiantes de la muestra, podremos estimar mejor el valor de μ (Agresti, 2018, pp. 132-138). Los resultados muestran que el *IC* calculado para la generación 2013 no se superpone con el *IC*

calculado para las generaciones 2012-2018 (tabla 1), lo que indica que los elementos que componen la muestra de la generación 2013 no son suficientes para representar la diversidad de la población. Podemos emplear también el intervalo de confianza del valor medio para comparar poblaciones en base a datos recabados de muestras concretas. La tabla 1 lista los promedios (\bar{x}) de los puntajes y el *IC* del valor medio (μ) de la población para las muestras de estudiantes de la generación 2013 que cursaron Química General I asistiendo a clase en forma presencial ($n = 442$) o mediante la utilización de materiales didácticos con guías de resolución de ejercicios disponibles en línea ($n = 36$). El promedio del puntaje de los estudiantes presenciales (18,5) es mayor que el de los que cursan en línea (16,6), pero ¿esto significa que el estudiante presencial tiene un puntaje medio mayor que el que cursa a través de la web? Para saberlo, hay que comparar los valores de μ de ambas poblaciones, estimados a partir de sus intervalos de confianza. En este caso, los *IC* de los valores de μ se solapan, por lo que no podemos concluir (a partir de estos datos), con un 95 % de confianza, que al estudiante presencial promedio (μ está en el rango 17,9 – 19,1) le vaya académicamente mejor que al estudiante promedio que cursa en línea (μ está en el rango 15,1 – 18,1).

Para poder comparar mejor los puntajes de las poblaciones de estudiantes presencial y en línea, necesitamos mejorar las estimaciones de ambos valores de μ (Agresti, 2018, pp. 132-138). Esto se logra, como hemos visto, incrementando el número de estudiantes en la muestra, lo que agrega información al modelo estadístico. En especial para los estudiantes que cursan en línea, que es un subgrupo con un número mucho menor de estudiantes, el emplear una muestra más grande mejorará sensiblemente la estimación que podamos hacer del valor medio del puntaje. Para esto, tomemos nuevamente como muestra los estudiantes presenciales y en línea, pero de las generaciones de 2012 a 2018 (incluyendo la totalidad de los estudiantes que cursaron en ese período, $n_{total} = 3.746$). Ahora las distribuciones de los promedios de estudiantes presenciales y en línea son mucho más estrechas y los *IC* no se solapan (tabla 1). Podemos concluir, con un 95 % de confianza, que al estudiante presencial promedio (μ está en el rango 19,0 – 19,4) le va mejor académicamente que al estudiante promedio que cursa en línea (μ está en el rango 16,2 – 17,8).

Test de hipótesis como herramienta eficiente de comparación

Otra de las herramientas de la estadística inferencial es el *test de hipótesis* (Agresti, 2018, pp. 191-202). En este se hace una suposición sobre la población (*hipótesis nula*: ausencia de efecto) y luego se trata de verificar, empleando los datos de la muestra, si se posee suficiente evidencia para rechazar o no esa hipótesis frente a una *hipótesis alternativa* que nos interesa. En nuestro caso hemos elegido la hipótesis nula (H_0) de que los valores de μ de los estudiantes presenciales y en línea son iguales ($H_0: \mu_{pres} = \mu_{línea}$), es decir, que la modalidad en la que cursan los estudiantes no tiene efecto sobre su desempeño académico medio. Por otra parte, la hipótesis alternativa (H_a) podría ser que $\mu_{pres} > \mu_{línea}$. Para comparar las muestras de estudiantes presenciales y en línea del año 2013 e inferir qué ocurre con los correspondientes valores de μ , se debe calcular el estadístico t de la siguiente manera: $t = (\bar{x}_{pres} - \bar{x}_{línea})/s_e$, donde $s_e = \sqrt{\frac{s_{pres}^2}{n_{pres}} + \frac{s_{línea}^2}{n_{línea}}}$. Cuanto mayor es el valor de t , mayor es la diferencia entre los promedios de las dos muestras, y por tanto mayor es la evidencia de que ambos valores de μ son diferentes. Tomando los valores de la tabla 1, el resultado es $t = 2,29$. El siguiente paso es calcular la probabilidad (P) de que, si H_0 es verdadera, t sea mayor o igual a ese valor.ⁱⁱ El valor de P resultante en nuestro caso es 0,01. Por tanto, existe solo un 1 % de probabilidad de que, siendo H_0 válida, t valga 2,29 o más. En este sentido, sería muy raro que la hipótesis nula fuera verdadera, por lo que tenemos suficiente evidencia como para rechazarla en favor de H_a . Concluimos entonces, con más de un 95 % de confianza, que $\mu_{pres} > \mu_{línea}$, es decir, que al estudiante presencial promedio le va mejor que al estudiante promedio que cursa en línea. Es interesante notar que, como vimos antes, recurriendo a los *IC* de la generación 2013 no pudimos

concluir a este respecto, ya que se solapaban. Cuando esto ocurre, los tests de hipótesis son herramientas inferenciales mucho más eficientes para comparar dos poblaciones. Si aplicamos un test de hipótesis similar, pero empleando la muestra 2012-2018, llegamos a la misma conclusión ($t = 5,51, P = 4,1 \times 10^{-8}$).

Evaluación de la intervención didáctica: intervalo de confianza y test de hipótesis

Emplearemos ahora algunas de las herramientas discutidas para inferir acerca del efecto de la intervención realizada. Si comparamos el puntaje del estudiante promedio en base a muestras de los años previos a la intervención (2012-2016; $n = 2.661$) con el mismo parámetro durante el período de la intervención (2017-2018; $n = 1.086$), vemos que los IC correspondientes (tabla 1) no se solapan. Podemos concluir, con un 95 % de confianza, que el desempeño académico del estudiante promedio antes de la intervención (μ entre 18,6 y 19,2) fue menor que luego de esta (μ entre 19,2 y 20,0). Podemos además verificar esta conclusión realizando el test de hipótesis: $H_0: \mu_{2012-2016} = \mu_{2017-2018}$ y $H_a: \mu_{2012-2016} < \mu_{2017-2018}$. En este caso $t = 2,90$ y $P = 0,002$. Por tanto, existe una probabilidad muy baja (0,2 %) de que, siendo H_0 verdadera, el estadístico t dé un número tan grande. Tenemos suficiente evidencia, por tanto, para rechazar H_0 y concluir que el puntaje del estudiante de primer año promedio aumenta significativamente luego de la intervención.

La información obtenida a partir del análisis inferencial puede ser complementada para cada muestra con el uso de herramientas estadísticas descriptivas (Veiga, Otero y Torres, 2018) como los diagramas de caja de la figura 3. Podemos concluir que, además de aumentar el puntaje promedio, la intervención logró disminuir el número de estudiantes con puntajes muy bajos, a la vez que favoreció que el tercer cuartil incrementara su desempeño académico.

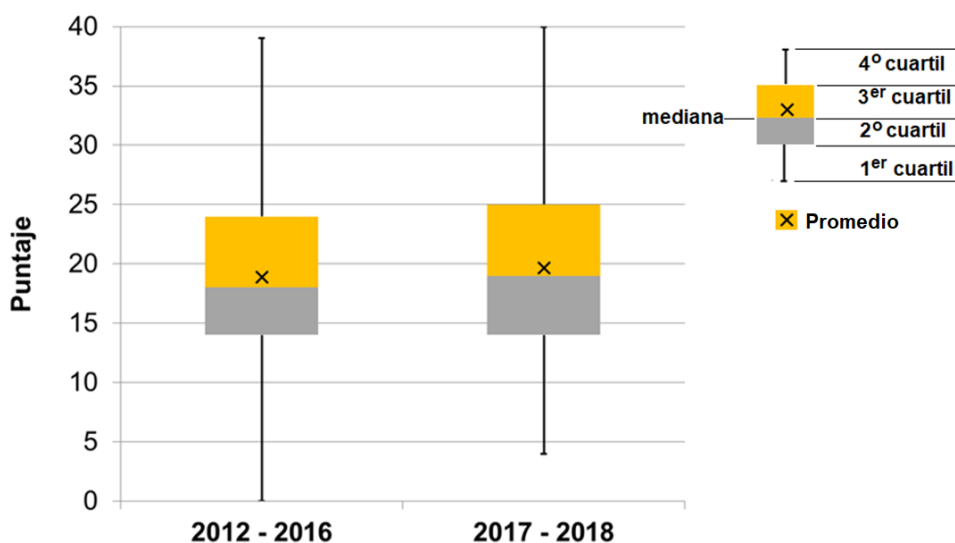


Figura 3: Diagramas de caja del puntaje de los estudiantes de las generaciones 2012-2016 y 2017-2018. Se representan los cuartiles que dividen la muestra de estudiantes en cuatro partes iguales, ordenadas de acuerdo al puntaje obtenido.

Conclusiones y limitaciones

Las herramientas estadísticas inferenciales presentadas (distribución de probabilidad, intervalo de confianza y test de hipótesis) permiten realizar predicciones, con un cierto nivel de confianza, sobre

cómo se comporta la población en estudio, a partir de los datos de muestras de la población. Dado que estos instrumentos permiten además comparar poblaciones, encuentran aplicación en la evaluación de la eficiencia de modalidades de trabajo e intervenciones didácticas.

En particular para nuestro caso de estudio, el análisis inferencial presentado permitió concluir, con un 95 % de confianza, que la modalidad de cursado (presencial o en línea), así como el uso de nuevos materiales de estudio interactivos influyen en el desempeño académico de los estudiantes de primer año de la Facultad de Química de la Udelar. Estos resultados son un insumo muy importante a la hora de la toma de decisiones académicas en nuestro departamento docente.

En primer lugar, una de las razones que podrían contribuir a la asociación entre la modalidad de cursado y el desempeño académico es que los estudiantes que asisten a clases presenciales tienen una mayor interacción con el docente y con sus pares, intercambian ideas y potencian así el proceso de aprendizaje. Además, la modalidad en línea incluye la realización de actividades fuera de clase, lo que requiere un mayor grado de disciplina y autorregulación para lograr un buen desempeño (Lai y Hwang, 2016). De todas maneras, existen en este sentido aspectos que no fueron controlados y que podrían estar influyendo en los resultados, ya que en la población de estudiantes en línea se observa un mayor porcentaje de personas que trabajan y también de personas que recursan. Estos aspectos serán sujeto de investigaciones futuras.

En segundo lugar, en las actuales condiciones de masividad en las que estos cursos se dictan no es posible atender en forma personalizada a cada estudiante, lo que desfavorece a los estudiantes con menores conocimientos previos y esto exacerba las diferencias existentes al llegar a la universidad. Los resultados de este trabajo implican en este sentido que el apoyo a las instancias de estudio mediante la inclusión de los nuevos materiales digitales interactivos puede constituir un aporte muy relevante para la mejora de los desempeños académicos por medio de la promoción de la autorregulación (Gravina y Prieto, 2019). En nuestro caso y en base a los resultados obtenidos, los materiales de estudio interactivos se incluyeron de forma permanente en los recursos didáctico-bibliográficos que brinda el curso. No obstante, tal como sucede con relación a la modalidad de cursado, para poder confirmar que existe una relación directa de causalidad sería necesaria la inclusión en el estudio de otras posibles variables (factores socioculturales, coyunturales, académicos, etc.) que también puedan influir en los resultados (véase por ejemplo Veiga, Luzardo, Irving, Rodríguez-Ayán y Torres, 2019).

Un aspecto importante a tener en cuenta son las limitaciones que poseen las herramientas inferenciales, lo que debe tenerse presente a la hora de extraer conclusiones fiables. En primer término, para poder emplear el análisis inferencial que hemos visto en este trabajo, la población en estudio debe ser infinita o tener un número extremadamente grande de elementos, de forma que su distribución de probabilidad sea continua y se cumpla que $s_{\bar{x}} = s/\sqrt{n}$ (Agresti, 2018, p. 112). En nuestro caso de estudio esto se cumple en la medida en que la definición de población adoptada incluye a todos los estudiantes de primer año de la Facultad de Química, de forma genérica y atemporal. Incluso aunque la población fuera definida como el subconjunto contemporáneo de esos estudiantes (por ejemplo de las últimas tres décadas), las herramientas inferenciales podrían aplicarse con éxito, ya que la población sería muy grande y su estudio completo (censo) no sería factible, porque parte de la información no está disponible, las evaluaciones eran muy diferentes, etcétera.

Por otra parte, la prueba con la que se miden los puntajes debería ser la misma para todos los estudiantes de la muestra. Esto no es factible cuando tenemos varias generaciones, dado que poner la misma prueba iría en desmedro de las garantías de una evaluación efectiva. Sin embargo, en nuestro caso esta suposición se cumple razonablemente, ya que todas las evaluaciones son de carácter múltiple opción y se confeccionan empleando una base de datos de ejercicios de nivel similar.

Finalmente, para obtener conclusiones confiables, las herramientas inferenciales que vimos dependen estrictamente de un muestreo aleatorio de una gran cantidad de sujetos, de forma que todos los “tipos” de estudiantes de la población estén representados en las muestras. Es por esto que en nuestro caso fue necesario incluir más de una generación de estudiantes, lo que posibilitó

mejorar la representatividad de la muestra y obtener la mayor información posible de la población en estudio. Estas distintas generaciones de estudiantes estarán probablemente afectadas por situaciones particulares que influyan en la variable a estudiar, lo que incrementa la proporción de la diversidad de la población que es explicada por las muestras que involucran varias generaciones.

Referencias bibliográficas

- Agresti, A. (2018). *Statistical Methods for the Social Sciences* (5.^a ed.). Pearson Education.
- Arim, R., y Katzkowicz, N. (2017). Trayectoria estudiantil: Determinantes de la deserción y culminación del ciclo educativo de estudiantes universitarios. *InterCambios*, 4(2), 109-127. <https://doi.org/10.29156/v4.i2/11>
- Gamboa, M. E. (2018). Estadística aplicada a la investigación educativa. *Revista Dilemas Contemporáneos. Educación, Política y Valores*, V(2), artículo n.º 5. Recuperado de <https://www.dilemascontemporaneoseducacionpoliticayvalores.com/index.php/dilemas/article/view/427>
- Gravina, V., y Prieto, V. (2019). La brecha secundaria-universidad: Diagnóstico de los estudiantes de Agronomía en el curso inicial de Matemáticas. *InterCambios*, 6(1), 34-42. <https://ojs.intercambios.cse.udelar.edu.uy/index.php/ic/article/view/205>
- Imer, M., Álvarez, N., Luzardo, F., Pacheco, M., Martínez, L., Veiga, N., ... Torres, J. (2018). Aprendizaje activo basado en materiales interactivos, una experiencia en primer año de grado de química a nivel universitario. *Anuario Latinoamericano de Educación Química*, XXXIII, 99-104.
- Lai, C., y Hwang, G. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education*, 100, 126-140.
- Maguire, T., y Rowley, G. (1992). Roles for Inferential Statistics in Educational Research. *Educational Philosophy and Theory*, 24(2), 56-77. <https://doi.org/10.1111/j.1469-5812.1992.tb00225.x>
- Veiga, N., Otero, L., y Torres, J. (2018). Análisis descriptivo de datos como herramienta en investigación didáctica. *Anuario Latinoamericano de Educación Química*, XXXIII, 210-215.
- Veiga, N., Luzardo, F., Irving, K., Rodríguez-Ayán, M. N., y Torres, J. (2019). Online pre-laboratory tools for first-year undergraduate chemistry course in Uruguay: Student preferences and implications on student performance. *Chem. Educ. Res. Pract.*, 20, 229-245. <https://pubs.rsc.org/en/content/articlelanding/2019/rp/c8rp00204e#!divAbstract>
- Zaiontz, C. (2020). *Real Statistics Using Excel*. www.real-statistics.com

Contribución de autoría

Los tres autores participaron por igual en las instancias de elaboración del artículo.

ⁱ <http://www.real-statistics.com/>

ⁱⁱ Esto puede hacerse por ejemplo mediante programas en línea gratuitos, como el disponible en https://istats.shinyapps.io/2sample_mean/ o mediante complementos sin cargo para Excel (Zaiontz, 2020).