

Evaluación de los Índices de Reprobación de la Universidad usando Intervalos de Confianza

Evaluation of University's Failure Indexes using Confidence Intervals

Héctor Francisco Ponce Renova

Universidad Autónoma de Ciudad Juárez/ Doctorado en Investigación Educativa

hector.ponce@uacj.mx

Resumen

El objetivo del presente estudio es evaluar los índices de reprobación del segundo semestre del 2014 en una universidad del Norte de México al igual que proponer una nueva metodología para analizar datos en las ciencias sociales. Esta nueva metodología plantea usar intervalos de confianza con la mediana para la detección de casos atípicos/extremos (*outliers*) porque es menos sensible que el promedio a datos sesgados. Se incluyen ejemplos de cómo calcular intervalos de confianza con la mediana. Los datos incluyen todas las calificaciones de estudiantes ($n = 130,452$). La estructura de los datos fue: cinco institutos, 16 departamentos y 123 carreras (i.e., 63 pregrados y 41 pregrados). Los resultados mostraron que existen casos atípicos a nivel facultad (un caso); departamentos (dos casos); y pregrados (tres casos). Estos resultados podrían ayudar a los hacedores de decisiones a dirigir sus esfuerzos para tratar de disminuir el número de calificaciones reprobatorias. Más allá de los resultados en particular, la nueva metodología no solo podría aplicarse en los índices de reprobación sino también en deserción, egreso, satisfacción de estudiantes con la universidad, etc.

Palabras clave

Estadísticas educativas, evaluaciones institucional, investigación aplicada, intervalos de confianza.

Abstract

The objective of this paper is to evaluate the failing index of the 2014s second semester in a northern Mexican university, as well to propose a new methodology to analyze data in social sciences. This new methodology presents the use of confidence intervals with the median to detect atypical/extreme cases because it is less sensitive to biased data. Examples included showing how to calculate confidence intervals using the median. The data included all the student's grades ($n = 130,452$). The backbone of the data is comprised of five institutes, 16 departments and 123 degrees (63 undergraduates and 41 graduates). The results showed atypical cases at an undergraduate level (one case); departments (two cases), graduate (three cases). These results could help the decision makers to channel their efforts into diminishing the number of failing grades. Beyond the results,

the new methodology could not only be applied in failing indexes but into deserting, graduating, student satisfaction, etc.

Keywords

Educational statistics, institutional evaluation, applied research, confidence intervals.

Introducción

Las universidades se esfuerzan por servir a su población estudiantil en la medida de sus posibilidades, produciendo exitosos y buenos ciudadanos del mundo (Thorne y Williams, 2012). Desde la adopción de la calidad total en los sectores públicos y privados, se le ha dado atención al usuario de un producto o un servicio (Sallis, 2014). Este servicio y atención a sus estudiantes enfrentan un problema: la reprobación. La reprobación de los estudiantes impide que algunos finalicen sus estudios, y, como consecuencia, no puedan contribuir como profesionistas en diferentes áreas (e.g., ingeniería, medicina, educación, etc.).

Ante esta problemática y usando una nueva metodología, el presente estudio plantea el objetivo de entender mejor el fenómeno de la reprobación y su asociación con otras variables (e.g., facultades, departamentos académicos, carreras de pregrado) al evaluar sus niveles en una universidad fronteriza en el segundo semestre del 2014. Este entendimiento podría llevar a implementar políticas educativas con el fin de disminuir la reprobación. Asimismo, la nueva metodología para analizar los datos de la reprobación por medio de intervalos de confianza con la mediana para identificar casos atípicos/extremos (*outliers*) podría ser usada en otros ámbitos de la investigación educativa. Para acercarse a este objetivo dual, se plantean la siguiente pregunta de investigación con su variable dependiente (reprobación) y variable independiente (i.e., facultades, departamentos y carreras de pregrado): ¿Existen índices de reprobación

atípicos/extremos (*outliers*)? Los índices de reprobación es una razón entre el número de personas que no pasaron una clase y el total que la tomo.

Los casos atípicos son valores poco probables que aparecen en los extremos de una distribución normal y pueden sesgar los promedios, afectar las comparaciones entre grupos y correlaciones (Osborne y Overbay, 2004; Kutner, Nachtsheim y Neter, 2004). Por estas razones, los casos atípicos son un problema para los análisis estadísticos. Primero, se dará un repaso de la distribución normal la cual es un fenómeno de la naturaleza que se da en la estatura, peso, e inteligencia de las personas, entre muchas otras cosas (ver a Hinkle, Wiersma y Jurs, 2003). La Figura 1 muestra una distribución normal con sus valores estandarizados: 0 (promedio), 1SD (desviación estándar), 2SD, y 3SD. Un valor estandarizado es una observación individual menos el promedio dividido por la desviación estándar:

$$(x_i - \text{promedio}) \div SD.$$

La probabilidad (área de la distribución normal) de que una persona este entre -1SD y 1SD es de 68%; entre -2SD y 2SD = 95%; y entre -3SD y 3SD = 99.7%. También, a estos coeficientes como $|1SD|$ se les llama valores zeta (z). Asimismo, estas probabilidades corresponden respectivamente a los intervalos de confianza de 68%, 95% y 99.7%. Un intervalo de confianza del 95% quiere decir que si tomamos 100 muestras de una población, esperaríamos que 95 de ellos capturarán el verdadero parámetro que buscamos y cinco no lo hicieron (Cumming, 2013).

Por ejemplo, supongamos que el promedio de altura de los varones de una muestra ($n = 100$) es de 1.80 m. con una SD de 10 cm. (ver Figura 1). También, en esta muestra se encuentra un hombre que mide 2.11 m. Para catalogar a un valor como atípico en una distribución normal (campana de Gauss; promedio = mediana = moda; y la curtosis y el sesgo = 0) tiene que estar más allá del valor absoluto de $|2SD|$ (4.30%), o más allá de $|3SD|$ (.26%) del promedio, otras cosas siendo iguales. Suponiendo que se crea un intervalo de confianza para este promedio de 1.80 m.:

- Promedio \pm Nivel de Confianza \times Error estándar
 - o el nivel de confianza puede ser de 95% ($2SD$) o del 99.7% ($3SD$).
 - o el error estándar (SE) es la desviación estándar de un set de muestras. Su fórmula es $SE = SD \div \sqrt{n}$. Donde n es el tamaño de la muestra.
- Para el ejemplo sería:
 - o $1.80 \pm 3 \times (10 \div \sqrt{100})$
 - Limite bajo del intervalo de confianza = 1.50 m.
 - Limite alto del intervalo de confianza = 2.10 m.

Como existe una persona de 2.11 m. en la muestra con un promedio de 1.80 y SD de 10 cm., él está a poco más de $3SD$ del promedio (fuera del límite alto del intervalo de confianza) y, por lo tanto, es un caso atípico/extremo con una probabilidad de suceder de 0.13%. Es decir, es poco probable que una persona de esa altura aparezca, dada la distribución estándar.

Además de la contribución del entendimiento de la reprobación, se planea una nueva forma de analizar los datos que es por medio de intervalos de confianza usando la mediana para detectar

casos atípicos (ver detalles de esta nueva metodología basado en las propuestas de Cumming, 2013, y Leys, Ley, Klein, Bernard y Licata, 2013, dentro del presente manuscrito). Aunque el método de intervalos de confianza con la mediana se había usado desde hace décadas en las ciencias exactas, parece ser que Leys et al. (2013) fueron los primeros en recomendarla para las ciencias sociales. La explicación sería que las poblaciones contienen parámetros y una muestra contiene estadísticas (estimaciones de los parámetros) en las cuales se asume que no corresponden totalmente a la población porque existe un error llamado error de muestreo (Cumming, 2013). Por ello, este último autor dijo que el manejar solo estadísticas sin considerar este error de muestreo sería un reporte incompleto ya que los límites bajo y alto de un intervalo de confianza son las estimaciones de este error entre la estadística de la muestra y el parámetro de la población.

Para evaluar la literatura de parte de la metodología del presente manuscrito en términos de reprobación e intervalos de confianza, se realizó una búsqueda en Google Académico durante el periodo de tiempo del 2014-2016. El periodo de tiempo para la revisión de la literatura fue posterior al de Cumming (2016) y Leys et al. (2013). Se concluyó que existe un vacío tanto en el idioma inglés como en el español porque al hacer las búsquedas en cada idioma no aparecen artículos que contengan todas las palabras clave (i.e., *college*, *confidence intervals*, *failure*, y *outliers*; intervalos de confianza, reprobación, universidad, valores atípicos).

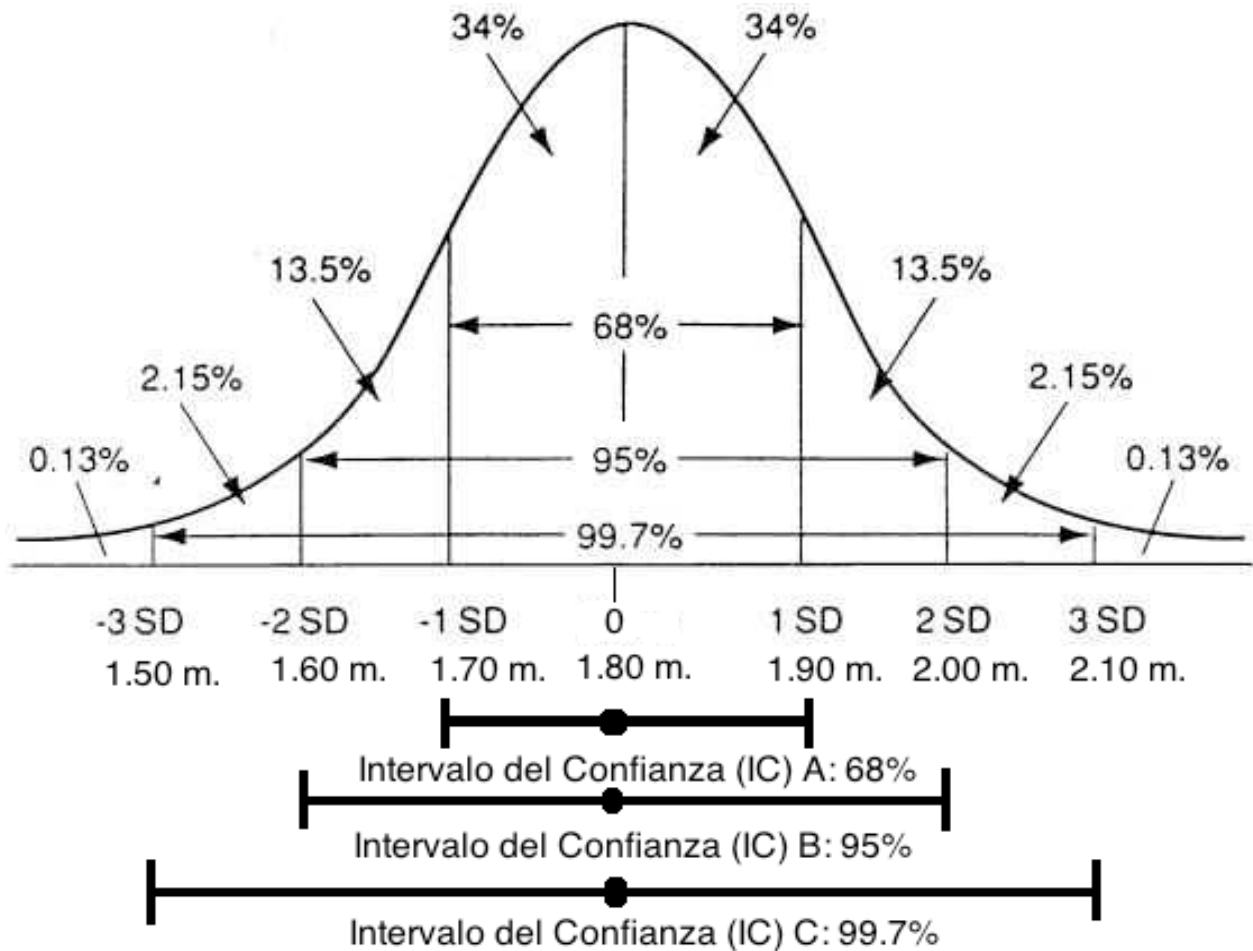


Figura 1. Distribución Normal e Intervalos de Confianza, con valores estandarizados sus probabilidades correspondientes y el ejemplo de la estatura de una muestra de varones.

Metodología

Las estadísticas descriptivas del presente estudio muestran que en una base de datos (la del segundo semestre del 2014) tuvo 130,452 observaciones, incluyendo los pregrados, maestrías y doctorados. Cada observación es una calificación de un estudiante durante ese semestre. En este periodo de tiempo, hubo aproximadamente 27,000 estudiantes en esta universidad. Esto hace que el promedio de calificaciones haya sido de aproximadamente de cinco calificaciones para cada estudiante en este semestre. Por confidencialidad se ocultaron los nombres de las facultades, los departamentos y las

carreras. Las facultades contuvieron el número correspondiente de departamento académico: en facultad I (1 departamento); facultad II (4); facultad III (4); facultad IV (4); y facultad V (3). En total se analizaron 104 carreras de las cuales 63 fueron de pregrado y 41 posgrados. La Tabla muestra el número de calificaciones por departamento y facultad. Las calificaciones fueron desde 297 para la facultad I hasta 51,888 para facultad III. Los índices de reprobación fueron calculados: número de personas con calificación menor a siete (i.e., reprobados) entre el número total de personas que tomo la clase.

Tabla
 Numero de calificaciones por facultad y sus departamentos correspondientes.

Departamento/ Facultad	I	II	III	IV	V
I	297				
II-A		17,398			
II-B		13,190			
II-C		3,633			
II-D		4,211			
III-A			9,498		
III-B			25,484		
III-C			11,093		
III-D			5,813		
IV-A				12,477	
IV-B				1,025	
IV-C				6,974	
IV-D				5,405	
V-A					7,875
V-B					3,993
V-C					2,086
Total	297	38,432	51,888	25,881	13,954

Para analizar los casos atípicos con intervalos de confianza del 99.7%, se utilizaron los *nuevos métodos* de análisis estadísticos para las ciencias sociales que propusieron Cummings (2013), y Leys et al. (2013). Por una parte, Cumming (2013) propuso que los intervalos de confianza sean utilizados en las ciencias sociales porque al usar una muestra se asume que hay un error entre un parámetro (e.g., un promedio de una población) y una estadística (e.g., el promedio de la muestra). Con el intervalo de confianza se estimaría con cierto grado de error (e.g., 5%) el parámetro. Para complementar los intervalos de confianza, Leys et al. (2013) propusieron el utilizar el valor de la mediana en lugar del promedio para detectar valores extremos. La razón fue que el promedio sería susceptible a los valores extremos (e.g., un valor muy alto como la fortuna de algún billonario puede

hacer que el promedio de la riqueza de una muestra suba sin que este promedio refleje con precisión cuanta riqueza posee la persona promedio). En cambio, la mediana no lo sería. Como parte de los análisis en el presente manuscrito, se utilizaron ambas propuestas de estos autores para detectar valores extremos (índices de reprobación que se salen del intervalo de confianza). Esta lista de autores fue seleccionada porque sus artículos y textos han respondido a preguntas de investigación similares a las del presente estudio.

Los intervalos de confianza con la mediana se calcularon de acuerdo a Leys et al. (2013) y el siguiente es un ejemplo heurístico para ilustrar:

1.- Primero hay que sacar la mediana de un set de números (ya sea de usando Excel, SPSS, etc.). Una manera manual con un valor atípico incluido sería ordenándolos: 2, 3, 4, 5, 1,000. La mediana = 4 (en cambio

el promedio hubiera sido 202.8; altamente susceptible a valores extremos como en este caso: 1,000).

2.- Para obtener lo que Leys et al. (2013) llaman: *Median Absolute Deviation (MAD;* Desviación Absoluta de la Mediana que es un análogo de la desviación estándar), se calcula así $|x_i - \text{Mediana}|$: $2 - 4 = |2|$; $3 - 4 = |1|$; $4 - 4 = |0|$; $5 - 4 = |1|$; $1,000 - 4 = |996|$. Ordenando los números: 0, 1, 1, 2, y 996. Luego se saca una mediana de esta diferencia = 1. Esta mediana se multiplica por la constante 1.4826 que es un valor ligado a la normalidad de los datos que excluye la anormalidad inducida por los valores extremos (ver a Rousseuw y Croux, 1993). Es decir, el *MAD* es: $1 \times 1.4826 = 1.4826$.

3.- Para establecer el intervalo de confianza, sería necesario definir si el investigador es lo que llama Miller (1991): muy conservador (*3SD*), moderadamente conservador (*2.5SD*) o no conservador (*2SD*). En otras palabras, si se elige el valor de *3SD*, el intervalo de confianza sería más grande y, por lo tanto, capturaría más valores extremos que si se hubiera usado el *2.5SD* o el *2SD*. Estos valores de *3SD*, *2SD* y *2.5SD* representan diferentes áreas bajo una distribución normal (figura 1). Para el presente manuscrito se eligió *3SD* para los análisis porque se quería detectar a los valores más atípicos, por decirlo de alguna manera.

-La parte inferior del intervalo de confianza se calcula: $\text{Mediana} - (3 \times$

$\text{MAD})$; $4 - (3 \times 1.4826) = 4 - 4.4478 = -.4478$.

- La parte superior del intervalo: $\text{Mediana} + (3 \times \text{MAD})$; $4 + (3 \times 1.4826) = 4 + 4.4478 = 8.4478$.

Los límites de este intervalo de confianza son: $-.4478$ a 8.4478 . En resumen, cualquier número que sea menor a $-.4478$ sería un valor atípico con un 99.7% de confianza. De una manera similar, un número como el 1,000, del ejemplo anterior, es un valor atípico con un 99.7% de confianza. El nivel de confianza sería en este caso que si obtenemos mil muestras de una población, esperaríamos que 997 de ellas capturaría el para metro de la población y tres de ellas no.

Resultados

Se calculó un intervalo de confianza del 99.7% para las facultades con sus índices de reprobación (número de reprobados en el instituto / número total el instituto). Estos resultados incluyeron los pregrados, maestrías y doctorados. Este intervalo de confianza iba del 6.3% (0.063) al 19.5% (.195). Es decir, sería normal tener índices de reprobación entre un 6.3% y 19.5% para las facultades, dados los datos. La figura 2 indica que solo la facultad IV (21.8%) se salió del intervalo de confianza. Esto indica que su índice de reprobación es extremo y, por ello, es más alto que los demás.

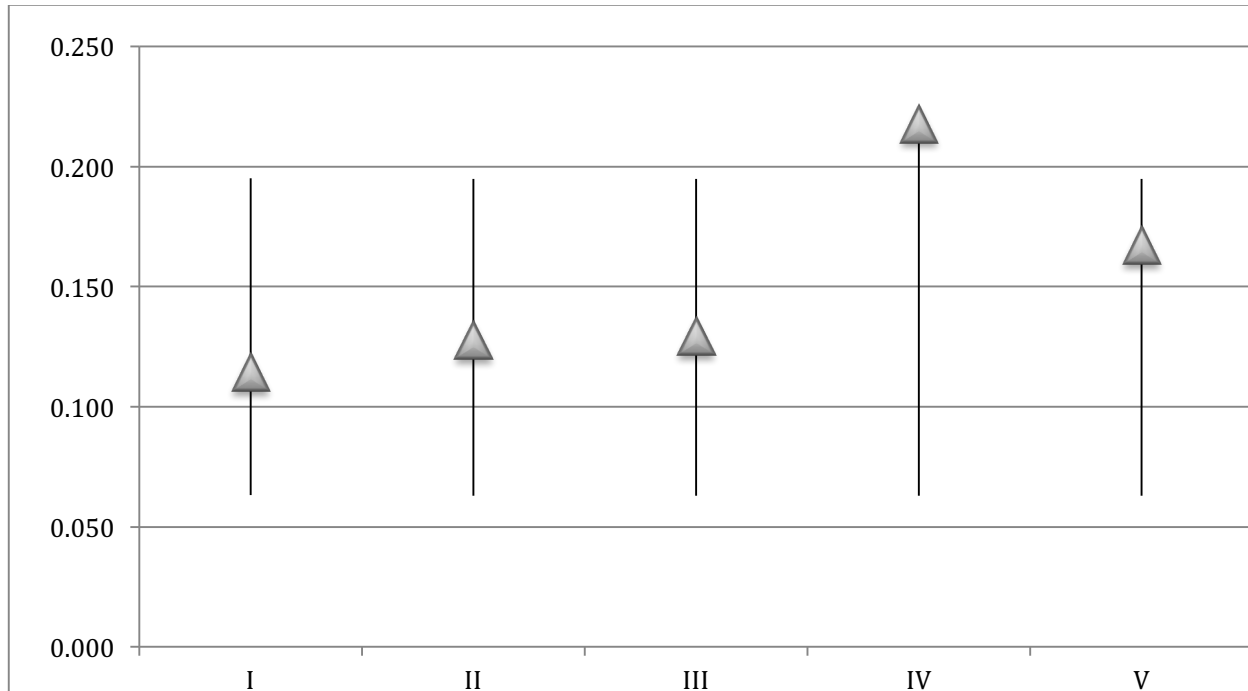


Figura 2. Intervalos de Confianza del 99.7% para las facultades.

Un análisis similar se hizo a nivel de los diferentes departamentos de cada facultad (ver figura 3). Estos resultados incluyeron los pregrados, maestrías y doctorados. Este intervalo de confianza del 99.7% estuvo entre 3.2% (0.032) al 25.2% (.252). Es decir, sería normal tener índices de

reprobación entre un 3.2% y 25.2% para los departamentos, dados los datos. Los departamentos IV-A (28.1) e IV-B (26.3) de la facultad IV tuvieron valores extremos de reprobación que el resto de ellos.

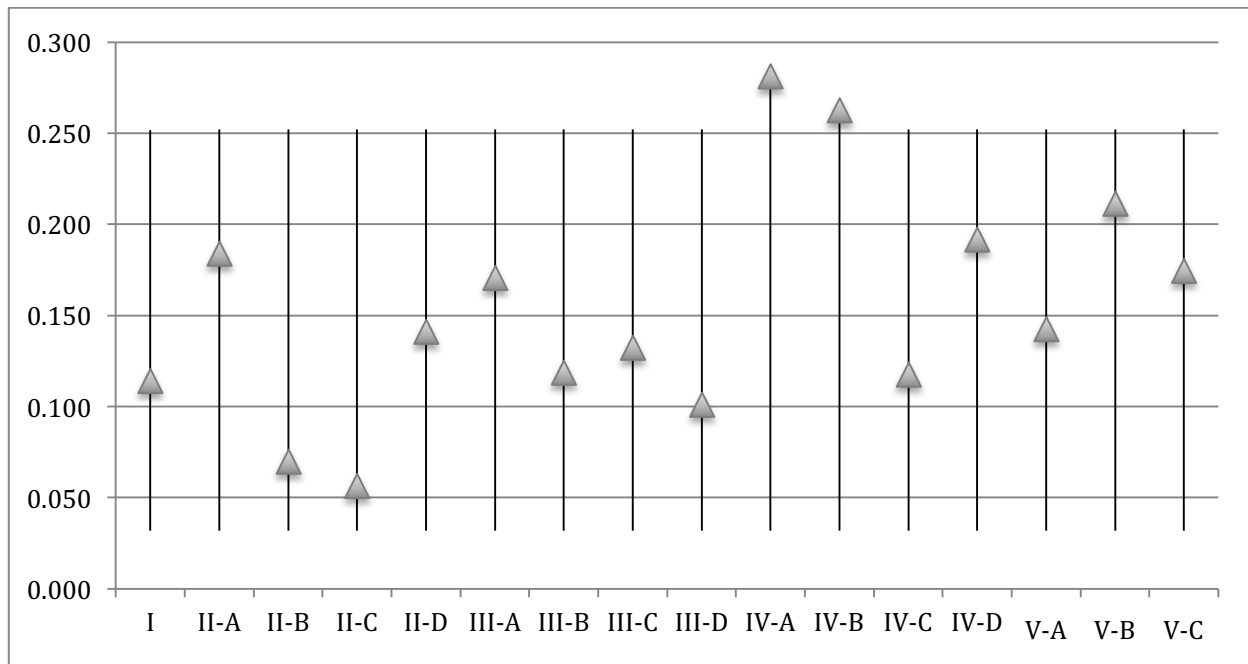


Figura 3. Intervalos de Confianza del 99.7% para los diferentes departamentos de las facultades.

La figura 4 muestra los resultados de los intervalos de confianza para las licenciaturas solamente. Las maestrías y doctorados no fueron incluidos por considerarlos en una categoría diferente debido a que el número de personas en cada postgrado difiere bastante del número de estudiantes en los pregrados. El intervalo de confianza del 99.7% para las

carreras de pregrado es 3.5% al 31.9%. Solo dos pregrados de IV-2 (34.6%) y IV-15 (39.2) tienen índices de reprobación extremos que superaron el límite alto del intervalo de confianza. Por otro lado, otra carrera de la facultad IV tuvo un índice más bajo que el límite bajo del intervalo de confianza: IV-16 (3.2%).

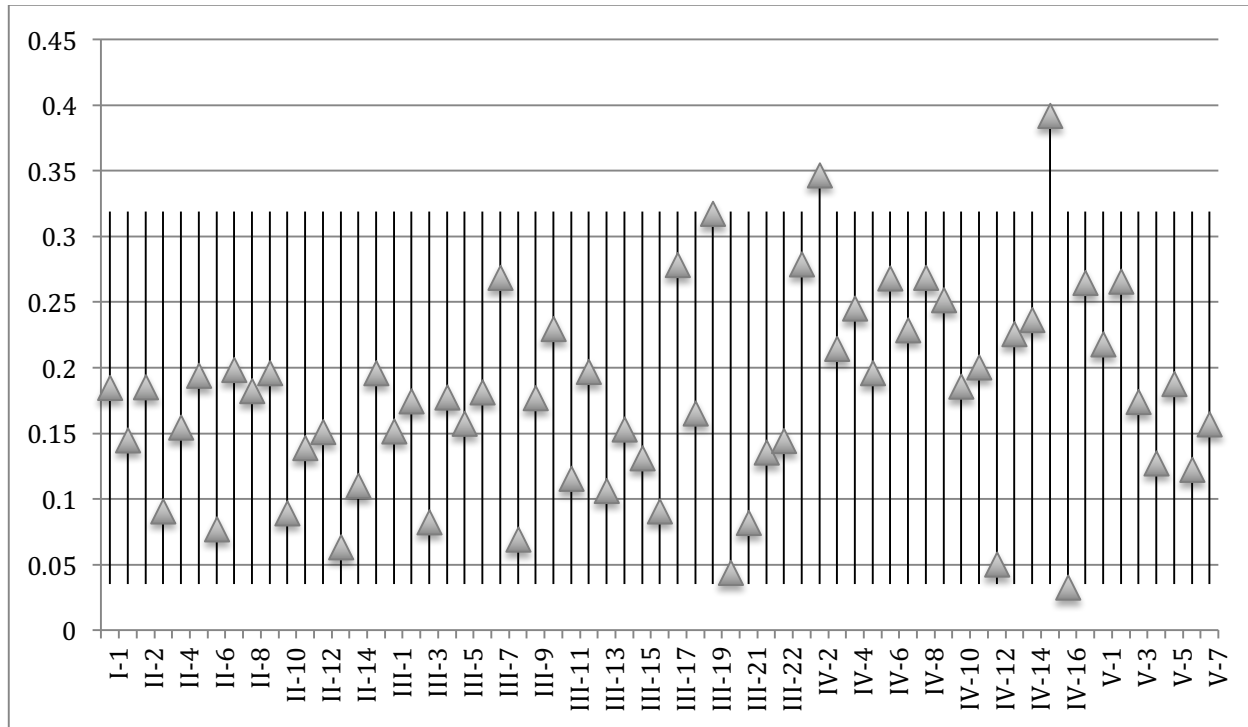


Figura 4. Intervalos de Confianza del 99.7% para algunas de las carreras de las facultades.

Conclusión y Discusión

Los análisis de casos atípicos fueron aplicados a la base de datos del segundo semestre del 2014 ($n = 130,452$ observaciones) a nivel facultad (5), departamentos (16) y carreras de pregrado (63). En general, un caso atípico es cuando un valor x se aleja de 2-3 desviaciones estándar del promedio (Cumming, 2013). Sin embargo, para este presente estudio se usara la sugerencia de Leys et al. (2013) quienes proponen usar la mediana en lugar del promedio porque esta primera estadística es menos sensible a los valores extremos. La pregunta de investigación del presente estudio fue: ¿Existen índices de reprobación atípicos/extremos (*outliers*)? La respuesta es: sí y estos son los resultados de los análisis que señalaron que si hay algunos casos atípicos entre las facultades (facultad IV), departamentos (IV-A y IV-B)

y carreras de pregrado (IV-2, IV-15 y IV-16; este último pregrado fue un caso atípico en el sentido que su índice de reprobación fue más bajo que el límite bajo). Es decir, los índices de reprobación excedieron los límites altos a estos tres niveles, así que tocaría el turno de los hacedores de decisiones investigar los posibles factores asociados a estos valores atípicos. También, se podrían implementar análisis de regresión múltiple para observar cómo se podrían asociar las calificaciones de los estudiantes con variables independientes como datos demográficos de estos, datos de sus profesores, etc. (para esta metodología ver a Tabachnick, 2012).

Entre las limitaciones del presente estudio están: (a) solo un análisis de una base de datos; y (b) el supuesto de que todos los estudiantes contaban con los recursos tangibles e intangibles, la capacidad intelectual y la motivación para

pasar sus respectivas clases. Esta limitación de una sola base de datos será resuelta en un futuro cercano cuando bases de datos sean incluidas y analizadas con intervalos de confianza también. De esta manera, se podrá observar si hay un patrón en los casos atípicos. Con respecto a la segunda limitación, habría que recopilar más datos de los estudiantes (e.g., trabajan, poseen computadora en casa, están motivados, etc.) por medio de encuestas y test para ver hasta donde se asocian estas variables con su nivel de calificaciones.

La otra contribución del presente estudio fue explicar las propuestas para las ciencias sociales de Cumming (2013) y Leys et al. (2013), con un par de ejemplos (ver la Introducción y Metodología para los detalles). El primero propone el uso de intervalos de confianza cuando se usan muestras porque estas contienen probablemente errores de muestreo. Las poblaciones contienen parámetros (e.g., estatura, inteligencia de personas, etc.) y las muestras contienen estadísticas (estimaciones de los parámetros). Con los intervalos de confianza se puede estimar este error (diferencia entre un parámetro y estadística). Un intervalo de confianza del 99% quiere decir que si se obtuvieran 100 muestras, 99 de ellas capturarían el parámetro de la población y una no. Para complementar los intervalos de confianza, Leys et al. (2013) propusieron que para evitar que una distribución normal sea sesgada por valores atípicos se use la mediana en lugar del promedio. La nueva metodología de estos últimos autores no solo ayuda a evitar el sesgo en el promedio sino también ayuda a detectar valores atípicos. En síntesis, se recomienda para la investigación educativa el uso de intervalos de confianza con la mediana para detectar valores atípicos que podrían

ser índices de reprobación, deserción, egreso, satisfacción de estudiantes con la universidad, etc.

Referencias

- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Nueva York: Routledge.
- Hinkle, D. E., Wiersma, W., y Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5ª ed.). Nueva York: Houghton Mifflin.
- Leys, C., Ley, C., Klein, O., Bernard, P., y Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764-766.
- Osborne, J. W., y Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(6).
- Kutner, M. H., Nachtsheim, C. J., y Neter, J. (2004). *Applied Linear Regression Models* (4ª ed.). Nueva York: McGraw Hill.
- Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424), 1273-1283.
- Sallis, E. (2014). *Total quality management in education* (3ª ed.). Sterling, Virginia: Routledge.
- Stice, J. D., Stice, E. K., Albrecht, W. S., y Skousen, K. F. (1999). *Survey of Accounting Principles*. Cincinnati, Ohio: South-Western College Publishing.
- Tabachnick, B. G., Fidell, L. S. (2012). *Using Multivariate Statistics* (6ª ed.). Nueva York: Pearson.

Thorne, F., y Williams, K. (2012). Librarians collaborate with the Office of Multicultural Affairs, a division of Student Affairs. *Library Services for Multicultural Patrons: Strategies to Encourage Library Use*, 49-52.

