





CONSISTENCIA INTERNA DE LA EVALUACIÓN ESTUDIANTIL DE DOS CUATRIMESTRES EN LA CARRERA DE INGENIERÍA EN PRODUCCIÓN INDUSTRIAL DE LA UNIVERSIDAD TÉCNICA NACIONAL EN SU SEDE CENTRAL ALAJUELA

José Mauricio Alcázar Román

RESUMEN

La consistencia interna es una evaluación estadística de qué tan seguro es que los elementos de las encuestas o pruebas que se diseñan para medir el mismo constructo (en este caso la evaluación realizada por medio de los estudiantes del turno de la tarde y de la noche, de los cursos ofrecidos durante el tercer cuatrimestre 2019 y el primer cuatrimestre 2021) de la carrera de Ingeniería en Producción Industrial en la Universidad Técnica Nacional, en su Sede Central en Alajuela, Costa Rica.

Por consiguiente, un alto grado de consistencia interna indica que los elementos diseñados para evaluar el mismo constructo generan puntuaciones similares; con ello se pone a prueba las evaluaciones de la carrera ante la escala utilizada y analizar con el indicador estadístico alfa de Cronbach (parámetro estadístico que mide la fiabilidad de consistencia interna), para determinar si las preguntas miden consistentemente la satisfacción del estudiantado en los periodos seleccionados.

Palabras claves: Alfa de Cronbach, constructo, nivel de satisfacción, fiabilidad, Minitab.

ABSTRACT

Internal consistency is a statistical evaluation of how accurate are the survey elements or tests designed for measuring a construct (in this case, a satisfaction survey of the courses offered by the school and responded by day and night-shift students during the third and first terms of 2019 and 2021 respectively) from the Industrial Production Engineering major in the Universidad Técnica Nacional's (UTN) main campus in Alajuela, Costa Rica.

Therefore, a high degree of internal consistency indicates that items designed to assess the same construct generate similar scores; with this, the evaluations of the career are put to the test before the scale used and analyzed with the statistical indicator Cronbach's alpha (statistical parameter that measures the reliability of internal consistency), to determine if the questions consistently measure the satisfaction of the student body in the periods selected.

Keywords: Cronbach alpha, construct, satisfaction level, reliability, Minitab.

José Mauricio Alcázar Román es Profesor en LEAD University en la Ingeniería en Productividad Empresarial y profesor en Universidad Técnica Nacional.

INTRODUCCIÓN

Es a través de instrumentos como las encuestas (evaluaciones o pruebas) que podemos medir los elementos de consistencia interna de los datos, y con ello establecer un marco referencial como primer paso a la mejora continua; un constructo es un tema, característica o habilidad subyacente, como la comprensión lectora o la satisfacción del cliente. Se busca que los elementos diseñados para medir el mismo constructo generen puntuaciones similares y con ello, lograr un alto grado de consistencia interna. Existe una variedad de medidas de consistencia interna, por lo general, implican determinar qué tan correlacionados están estos elementos y qué tan bien se predicen entre sí. El alfa de Cronbach es un indicador que se usa comúnmente para utilizar medidas de consistencia interna, los elementos por lo general deben estar en un solo instrumento de medición y ser administrados a un grupo de personas en una ocasión, con el fin de evitar variables de confusión (Minitab 1, 2021).

Para el artículo se utilizaron las evaluaciones realizadas para otorgarle fiabilidad estadística, por el estudiantado de la carrera de Ingeniería en Producción Industrial (IPRI) en su sede central en Alajuela (la carrera se ofrece también en la sede del Pacífico - Puntarenas), en dos periodos distintos (tercer cuatrimestre 2019 y primer cuatrimestre 2021) y en dos segmentos del día: horario de la tarde y de la noche; esto porque el horario de la mañana presenta muy pocos datos en ambos periodos y se recomiendan más de 30 datos para tener resultados consistentes.

REFLEXIÓN

En 1967 Nunnally (Maese, 2016) definía la fiabilidad como el grado en que las mediciones de los constructos son repetibles y cualquier influencia aleatoria que pudiese hacer los datos diferentes de una medición a otra es una fuente de error del valor obtenido. Por lo tanto, la fiabilidad indica una consistencia de la medida del constructo.

En este sentido, es deseable que se obtengan puntuaciones similares (consistentes) cuando el instrumento de medida se administra repetidamente a los sujetos bajo las mismas condiciones. Se trataría de un instrumento con una alta fiabilidad, es decir libre de sesgos. Si el valor de fiabilidad es bajo indicará que hay errores de medición y por lo tanto los resultados de los

análisis estadísticos estarían sesgados al carecer de fiabilidad o consistencia interna. Aquí se establece que se tienen variables de confusión, como por ejemplo preguntas mal formuladas que llevan a estudiantado a sesgar las respuestas basados en la comprensión de lo que se quiere evaluar.

Una definición formal de la fiabilidad desde el modelo de la teoría clásica de evaluación la define como la razón entre la varianza de la puntuación verdadera y la varianza de la puntuación total. (Cronbach, 1951). Lo anterior se puede comprender como el factor de variabilidad entre las respuestas, que se presumen están claramente identificadas y los resultados totales de la encuesta al estudiantado; generando un análisis estadístico paralelo al instrumento, más allá de saber por el pensamiento o percepción de las preguntas, sino también como efecto lateral clave, si la encuesta cuenta con representatividad estadística, así como la correcta aleatoriedad, libre de sesgos amplios por inadecuada interpretación de las preguntas.

El gran reto y pregunta sobre la mesa, se establece en como medir lo intangible, en servicios, por ejemplo. La respuesta inicial primeramente es que todo puede y debe ser medido; en segundo lugar, el instrumento y/o escala de medición, para lograrlo. Con ello podemos llevar aspectos del tipo percepción, a elementos cuantitativos que pueden ser utilizados para manipular valores numéricos y obtener medidas claves para el análisis y la toma de decisiones. En el pasado, la carrera no ha contado con un estudio similar que busque validar el instrumento y, por ende, la fiabilidad de los resultados. La única evolución que se conoce es ajustar y/o cambiar preguntas, buscando mejorar el instrumento, pero sin una base estadística de representatividad y disminución de la subjetividad.

Este artículo fue desarrollado con dos grandes objetivos: primero aprobar el curso de Publicación de Artículos Científicos, ofrecido a los docentes por parte de la Universidad Técnica Nacional; y segundo, como base de aprendizaje para aplicar en la tesis de maestría científica del Instituto Centroamericano de Administración Pública ICAP, defendida en marzo del 2022, que lleva por título: *Evaluación de la calidad en la prestación de los servicios administrativos para las carreras de ingeniería de la Universidad Técnica Nacional - sede central; tomando como parámetro la guía metodológica del Ministerio de Planificación Nacional y Política Económica (MIDEPLAN)*; donde la guía del

Mideplan otorga únicamente los pasos a seguir para la evaluación de los intangibles, pero no profundiza en elementos estadísticos propios de la valoración cuantitativa de los instrumentos posibles a utilizar.

Escala de Likert

Una escala típica usada para medir la información ordinal es la escala desarrollada por el investigador Rensis Likert, quien publicó en 1932 un informe donde describía su uso. La información ordinal es definida como información categórica con un orden lógico de las clasificaciones o dimensiones establecidas.

La clasificación de Likert es comúnmente usada como una escala psicométrica estándar para medir respuestas. Esta escala de medición tiene un procedimiento que facilita la construcción y administración de un cuestionario, así como la codificación y análisis de la información recabada en la revista de Maese (2016). Para un escalafón de cinco puntos, por ejemplo, cada uno podría ser etiquetado acorde a su nivel de preferencia: 1 = fuertemente en desacuerdo (FD), 2 = desacuerdo (D), 3 = ni acuerdo ni desacuerdo (NN), 4 = de acuerdo (A) y 5 = fuertemente de acuerdo (FA). Dependiendo de lo que se esté midiendo, las etiquetas de la clasificación pueden ser expresadas diferentemente.

Basado en análisis de la literatura y experiencia propia en este tipo de instrumentos, las recomendaciones para las escalas de Likert a utilizar son: manejar escalas no muy cortas (uno a tres por ejemplo), así como escalas muy largas (uno a diez); de utilizarse la última, cada ítem de la escala debe estar claramente definido para disminuir al máximo el sesgo de selección entre valores.

Coefficiente alfa de Cronbach

La fiabilidad se refiere a la medida en que un experimento, prueba o cualquier procedimiento de medición asegura replicar resultados bajo condiciones de validación similares. La fiabilidad de las puntuaciones es el grado en que los datos de una muestra particular están libres de error de medición aleatorio (importante replicar que el sesgo estadístico es inherente en todo proceso de análisis muestral y lo que se busca es que sea lo más bajo posible, para mejorar el nivel de precisión del análisis).

Autores confirman que el valor mínimo del coeficiente de fiabilidad debe ser 0,7, el valor más cercano a 1 es el índice de la mayor fiabilidad del rango. Escalas

entre 0,6 y 0,7 se consideran como notas mínimas de aceptación; pero se espera valores superiores a ellas; con niveles menores a 0,6 colocan al procedimiento de medición con una fiabilidad no adecuada y requieren de ajustes estructurales significativos, según Maese (2016).

El tipo de coeficiente de fiabilidad mayormente reportado en la literatura es el coeficiente llamado alfa de Cronbach. Este parámetro estadístico mide la fiabilidad de consistencia interna, grado en que las respuestas son consistentes a través de los ítems dentro de una medición. Si la consistencia interna es baja, entonces el contenido de los ítems puede ser tan heterogéneos que la puntuación total no es la mejor unidad posible de análisis para la medición. Conforme la consistencia interna se acerca a cero, las puntuaciones cada vez más y más se vuelven números aleatorios (no correlacionados entre sí) y los números aleatorios no miden nada, Maese (2016).

Análisis Multivariado con Minitab

Minitab es un software estadístico especializado para examinar datos actuales y del pasado, buscar tendencias, predecir patrones, descubrir relaciones ocultas entre las variables, visualizar las interacciones de los datos e identificar factores importantes para responder incluso a las preguntas y problemas más desafiantes. Las visualizaciones son algo bueno, pero al combinarlas con la analítica se convierten en algo extraordinario. Con el poder de las estadísticas y el análisis de datos de su lado, las posibilidades son infinitas (Minitab2, 2021).

Dentro de las herramientas de análisis que posee este software estadístico, se tienen herramientas de comprensión multivariado; y dentro de esta categoría se localiza el análisis de elementos: el cual evalúa la confiabilidad de múltiples elementos de un sondeo para medir la misma característica. Si las respuestas a estos valores son internamente consistentes, entonces existe evidencia de que los elementos miden el mismo constructo de satisfacción al cliente. Por el contrario, si los valores no están escritos adecuadamente, los clientes podrían interpretarlos de manera diferente y las respuestas variarán de un significado a otro (Minitab2, 2021).

Interpretado lo anterior, el artículo se esfuerza en establecer o posiblemente “quebrar” el método tradicional de realizar evaluaciones para medir aspectos de servicios en cualquier sistema productivo o de atención;

donde la mayor cantidad de las ocasiones, el único enfoque estadístico o análisis, se basa en determinar los tamaños de muestra. El principal aporte mencionado lo anterior, consiste en el desarrollo del parámetro estadístico, que busca medir el nivel de consistencia interna de las respuestas a toda la muestra evaluada.

DESARROLLO

La carrera de Ingeniería en Producción Industrial de la Universidad Técnica Nacional, en su Sede Central en Alajuela desarrolla evaluaciones del nivel de satisfacción entre el estudiantado todos los cuatrimestres, sobre los cursos ofertados y en los tres bloques de clase que se ofrecen: mañana, tarde y noche.

La oferta por las mañanas es baja y no permite una cantidad de datos considerable para el análisis y pone en riesgo la consistencia estadística de los datos; por lo cual se seleccionan valores de las evaluaciones de los cursos de la tarde y noche, durante dos cuatrimestres que se utilizan para desarrollar el indicador estadístico alfa de Cronbach; se seleccionan 60 datos al azar de cada uno de los turnos: dos turnos de tarde y dos de noche, en el tercer cuatrimestre 2019 y el primer cuatrimestre 2021 (ver Cuadro 1). Los datos obedecen a la consulta sobre el nivel de satisfacción con respecto al curso a mediados del periodo, con un enfoque principalmente del docente a cargo, y se establece con una escala de valoración de Likert que va del 1 como valor de satisfacción más bajo y el 10 como valor de satisfacción mayor. Anteriormente se mencionó sobre las escalas del 1 al 10, al son escalas que son recomendadas a utilizar si cada uno de los valores dentro de la escala, está completamente descrita; esto para evitar sesgos entre tantos valores. Aspecto clave a considerar y validar para modelos de evaluación intermedios o complejos.

CUADRO 1: DATOS OBTENIDOS DE LOS TURNOS DE AMBOS PERIODOS

Jornada	Tercer cuatrimestre 2019	Primer cuatrimestre 2021
mañana	6	9
tarde	100	60
noche	137	62

Fuente: Elaboración propia, (2022).

Una vez tabulado los valores y utilizando muestras de mismo valor ($n = 60$) para todos los periodos, se utiliza el software especializado Minitab para analizar

los datos a través del análisis multivariado y obtener resultados que permitan concluir respecto a las evaluaciones realizadas por el estudiantado.

La herramienta necesita que los valores sean iguales (mismo tamaño de muestra o valor n) en los periodos analizados, por ello se selecciona la cantidad de sesenta como valor común, y en los periodos que poseen más datos, se selecciona una muestra aleatoria de la misma cantidad de valores. Se detecta, según el Cuadro 2, una media similar en los cuatro períodos (media global de 7.775), con variabilidades altas por el tipo de valores utilizados (media de la desviación estándar de 2.452 para los cuatro periodos); lo cual obedece a la escala utilizada de uno a diez, donde se tiene alta variabilidad de las evaluaciones en los cursos por parte del estudiantado.

CUADRO 2: ESTADÍSTICAS (MEDIA Y DESVIACIÓN) DEL ANÁLISIS OFRECIDO POR MINITAB.

Estadísticas totales y de elementos			
Variable	Conteo Total	Media	Desv. Est.
tarde 2021	60	8.700	1.660
noche 2021	60	7.650	2.773
tarde 2019	60	6.683	2.990
noche 2019	60	8.067	2.414
Total	60	31.100	5.108

Fuente: Elaboración propia, (2022).

Con respecto al indicador alfa de Cronbach, se obtiene un valor de 0,054 global, siendo apenas un 5 % lo cual está muy por debajo del número mínimo esperado (0,6 – 0,7), para poder concluir que los valores de la evaluación de medio ciclo están midiendo la misma característica en las mismas condiciones, para el estudiantado analizado. Los indicadores altos y positivos de la correlación total y múltiple del cuadro 3, indican que todos los elementos tienen baja o nula correlación entre sí. La correlación total (conocida también como coeficiente o momento de Pearson) es un elemento numérico que va desde el -1 (fuerte correlación negativa), pasando por el cero (correlación nula) hasta el $+1$ (fuerte correlación positiva); en este caso los valores se desplazan cerca del cero indicando lo antes mencionado: no se puede establecer estadísticamente una relación lineal entre los turnos evaluados.

El indicador alfa de Cronbach no logra establecer por los resultados del análisis realizado, un nivel de uniformidad para los múltiples elementos de la evaluación que hayan sido uniformes y precisos a la hora de

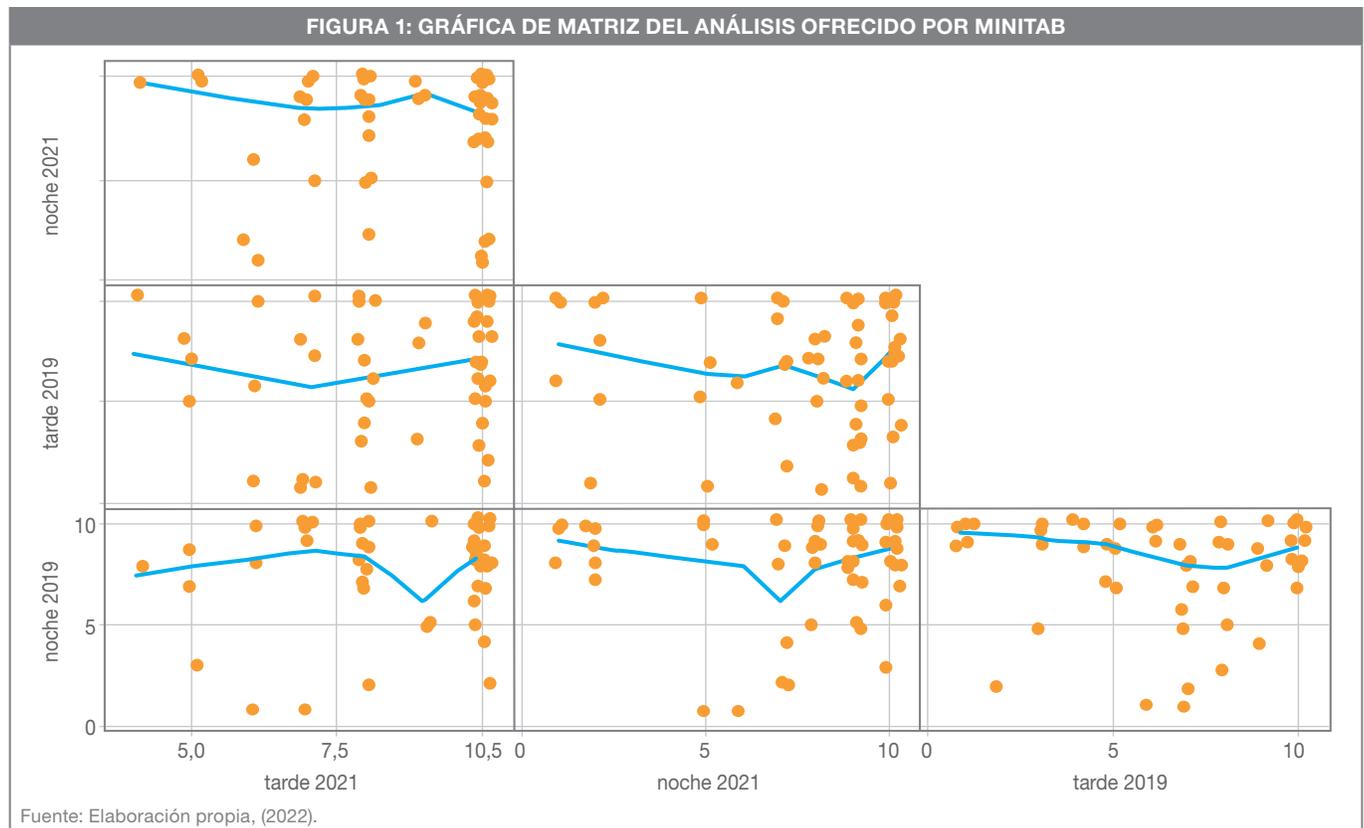
CUADRO 3: ESTADÍSTICAS (CORRELACIÓN Y OTROS) "DEL ANÁLISIS OFRECIDO POR MINITAB.					
Estadísticas de elementos omitidas					
Variable omitida	Media total ajustada	Desv. Est. total ajustada	Correlación total ajustada por elemento	Corelación múltiple cuadrada	Alfa de Cronback
tarde 2021	22.400	4.604	0,140163	0,057787	-0,076735
noche 2021	23.450	4.417	-0,045049	0,006181	0,166225
tarde 2019	24.417	4.163	-0,021992	0,021992	0,091826
noche 2019	23.033	4.380	0,039985	0,039985	-0,001723

Fuente: Elaboración propia, (2022).

evaluar la misma destreza o característica, en este caso el nivel de satisfacción de los estudiantes por los cursos ofertados; lo que llevaría a la carrera a realizar análisis individuales de cada curso, bloque (mañana, tarde o noche) y/o su nivel de satisfacción por separado.

Descriptivamente para comprender lo anterior, se presenta una gráfica de matriz realizada por Minitab, que evidencia lo analizado con las correlaciones y el indicador alfa de Cronbach. Las líneas rojas se establecen como rectas de mejor ajuste que buscan establecer correlaciones positivas o negativas, basadas en su pendiente. En todos los casos se observa las líneas bastante horizontales, lo que descarta una correlación positiva o negativa, entre las variables analizadas por el software.

Del Cuadro 3 otro aspecto a destacar son los valores de la desviación estándar total de cada una de las variables, siendo todas superiores al valor cuatro, lo cual es un valor muy alto y determina una fuerte dispersión de los datos. En este contexto se puede establecer como una variabilidad fuerte entre las evaluaciones del estudiantado, lo cual coloca como reto a la dirección de la carrera, buscar como alinear, estandarizar, normalizar los valores de medición para cada uno de los bloques futuros; o también, establecer una evaluación más estratificada entre los niveles educativos del estudiantado (no mezclar primer ingreso con estudiantes avanzados, por ejemplo).



En las distintas gráficas que evalúan la relación de los valores entre dos turnos (fuerza y dirección) para establecer la fuerza entre los ejes y buscar un impacto ya sea positivo, neutro o negativo de los datos, se observa la alta variabilidad de las evaluaciones obtenidas por el estudiantado. La gráfica de matriz destaca que los turnos de los respectivos períodos no están altamente correlacionados, entonces los elementos pudieran medir diferentes características o no estar claramente definidos para el estudiantado.

CONCLUSIONES

No se logra determinar el nivel de fiabilidad global del instrumento de evaluación para los cursos ofertados en los cuatro periodos analizados; considerando que lo anterior no desmerita la evaluación por parte de la dirección hacia los cursos en los bloques, lo cual busca el mejoramiento y un panorama basado en el nivel

de satisfacción por parte del estudiantado. Al observar los valores de media y desviación de los períodos, si se determina una consistencia entre las evaluaciones, lo cual justifica el instrumento.

Parte de la alta dispersión de los datos y, por ende, la afectación a la hora de analizar las distintas correlaciones entre los períodos se debe al utilizar una escala de Likert tan abierta con valores que van desde el uno como valor mínimo, hasta el diez como máximo; la investigación evidenció la utilización de escalas que en su mayoría utilizan valores del uno al cinco en los rangos, y cada uno bien definido para el estudiante a evaluar.

Los resultados obtenidos en el análisis no afectan o desacreditan las evaluaciones individuales realizadas por la dirección de la carrera cada período, que tienen por objetivo la obtención de información para los procesos de mejora establecidos por la Universidad, en miras de administrar, controlar y mejorar el nivel de los cursos impartidos cada cuatrimestre.

REFERENCIAS BIBLIOGRÁFICAS

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <http://doi.org/10.1007/BF02310555>
- Maese, J.; Alvarado, A. y Valles, D. (2016). Coeficiente alfa de Cronbach para medir la fiabilidad de un cuestionario difuso. *Revista Culcyt mexicana*, 13(59). <http://erevistas.uacj.mx/ojs/index.php/culcyt/article/view/1455>
- Soporte Minitab 1. (2021, marzo 29). *Interpretar los resultados clave para Análisis de elementos*. <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/statistical-modeling/multivariate/supporting-topics/item-analysis/what-is-internal-consistency/>
- Soporte Minitab 2. (2021, marzo 29). *Ejemplo de Análisis de elementos*. <https://www.minitab.com/es-mx/products/minitab/>