

Estadística y ciencia de datos para la evaluación en las intervenciones públicas

Este informe detalla los aspectos clave de un evento sobre la aplicación de la estadística y la ciencia de datos en la evaluación de intervenciones públicas, centrándose en la construcción y validación de un índice de bienestar universitario para medir el impacto de las becas.

1. Participantes e Instituciones Clave

El evento fue liderado por **Andrés Alejandro Galvis Correa**, un experto de la **ESPE (Escuela Politécnica del Ejército)**, quien compartió los avances de su institución en analítica, estadística y ciencia de datos, particularmente en la gestión del conocimiento y la gobernanza de datos.

Entre las instituciones y colaboradores mencionados durante la exposición se destacan:

- **ESPE (Universidad de las Fuerzas Armadas):** Anfitriona y principal presentadora del caso de estudio.
- **Secretaría Nacional de Planificación:** Mencionada como una entidad que debería definir parámetros y metodologías para la construcción de indicadores homogéneos.
- **INEC (Instituto Nacional de Estadística y Censos):** Al igual que la Secretaría Nacional de Planificación, se sugiere que el INE debería establecer metodologías estándar para la creación de índices.
- **Universidad de Princeton:** Referenciada a través del profesor **Matías Cattaneo**, una autoridad global en modelos de regresión discontinua, cuyos *papers* y paquetes son altamente recomendados.
- **FLACSO (Facultad Latinoamericana de Ciencias Sociales):** Se menciona a **Juan Ponce** como otro referente importante en evaluación de impacto en el contexto ecuatoriano.
- **Sociedad Ecuatoriana de Estadística:** Promocionada por Andrés Galvis como una entidad que ofrece cursos avanzados en estadística a nivel nacional y latinoamericano.
- **H.C.U. (Honorable Consejo Universitario):** Su rol se menciona en el establecimiento de los quintiles para el punto de corte de las notas que determinan la asignación de becas.

2. Temas Principales y su desarrollo

El evento se articuló en torno a la creación, validación y aplicación de un **Índice de Bienestar Universitario** para la evaluación de impacto. Los temas se desarrollaron de manera secuencial y práctica, utilizando un caso de estudio real:

2.1 La Estadística y Ciencia de Datos para la Evaluación de Intervenciones Públicas:

Es el uso riguroso de métodos cuantitativos para medir el impacto causal, la eficiencia y la efectividad de políticas, programas o acciones del Estado e implica a través de herramientas computacionales realizar lo siguiente:

- Realizar diseños experimentales o cuasiexperimentales (RCT, DiD, RDD, matching).
- Correr modelos estadísticos y de machine learning para predicción, segmentación y simulación.
- Analizar grandes volúmenes de datos administrativos, censales o de encuestas.
- Visualizar e interpretar los resultados para informar, tomar de decisiones basada en evidencia.

2.2 ¿Por qué evaluar intervenciones públicas?, las consecuencias de no evaluar y el reto del sector público, la causalidad, y la no correlación.

Entre sus principales razones por la que se evalúa son las siguientes:

- Garantizar eficiencia en el uso de recursos públicos.
- Asegurar la efectividad de las políticas implementadas.
- Evitar intervenciones ineficaces o contraproducentes.
- Tomar decisiones basadas en evidencia, no en intuiciones políticas o presiones externas.

Y las consecuencias de no evaluar las políticas públicas es la siguiente:

- Se mantienen programas sin impacto real.
- Se desperdician fondos públicos.
- Se toman decisiones basadas en percepciones, no en resultados.
- Se limita la capacidad de escalar soluciones exitosas o de corregir fallas.

Evaluar no es lo mismo que monitorear. Mientras el monitoreo describe como se planeó y qué ocurre", la evaluación responde al **"¿qué hubiese pasado si no hubiéramos intervenido?"** o ¿Cambio algo gracias a la intervención del estado? Para evaluar bien, se necesitan métodos rigurosos estadísticos que comparen escenarios (¿qué pasó con y sin la política), evitando atribuir efectos a simples coincidencias. En resumen: Evaluar es aprender, mejorar y rendir cuentas.

2.3 Ciclo de vida de Ciencia de Datos para Políticas Públicas

- Diagnóstico del problema público: Identificar las necesidades y definir los objetivos de política.
- Integración de fuentes de datos administrativas: Reunir datos relevantes provenientes de registros, censos y encuestas.
- Limpieza y preparación de datos: Corregir errores, tratar valores faltantes y estandarizar formatos.
- Análisis exploratorio: Explorar patrones, tendencias y formular hipótesis sobre el fenómeno de interés.
- Construcción de indicadores y dimensiones: Generar variables sintéticas, índices y tipologías que representen constructos clave.
- Modelamiento predictivo o causal: Entrenar modelos estadísticos o de machine learning para estimar impactos o realizar predicciones.

- Visualización e informe para la política: Comunicar hallazgos y recomendaciones mediante gráficos, dashboards o informes narrativos.
- Implementación y seguimiento del impacto: Desplegar soluciones, monitorear resultados y ajustar decisiones de política pública.

2.4 Estudio de Caso: Diseño y Validación de un Instrumento para Medir el Bienestar Universitario.

1. Contexto del problema:

La Universidad de las Fuerzas Armadas - ESPE busca diseñar un instrumento que permita medir el bienestar universitario de sus estudiantes con fines de análisis institucional, diagnóstico y evaluación de intervenciones.

Este instrumento pretende generar un **índice compuesto** que integre múltiples dimensiones del bienestar y permita:

- Evaluar impacto de programas (tutorías, becas, salud mental).
- Identificar grupos vulnerables.
- Tomar decisiones con base en evidencia cuantitativa.

Este taller simula el proceso completo de validación de dicho instrumento, integrando estadística multivariante, juicio de expertos y técnicas de ciencia de datos.

2. Objetivos del caso de estudio

- Construir un instrumento de medición del bienestar.
- Simular la validación de contenido (juicio de expertos).
- Simular la validación de criterio (entrevistas y prueba piloto).
- Aplicar modelos multivariantes (AFE, fiabilidad, validez).
- Calcular un índice compuesto de bienestar.

3. Metodología general

- Diseño del instrumento (dimensiones e ítems tipo Likert)
- Validación de contenido (índice de validez de contenido - IVC)
- Validación de criterio (entrevistas simuladas y prueba piloto)
- Validación estadística (AFE, alfa de Cronbach, validez)
- Construcción del índice final (ponderaciones y visualización)

4. Desarrollo del caso:

Para mayor detalle y como ejemplo del caso desarrollado en la ESPE, para la asignación de becas, se utilizó la herramienta en línea Google Colab, con la estructura de Python para la generación códigos. En el siguiente link de descarga se encuentra el desarrollo del estudio de caso.

https://colab.research.google.com/drive/1SSC3Jij980hwDMfwTnzMdUSODIVRP2_u#scrollTo=GBzY5LvHYmI6

- En el caso de estudio se utilizó como herramienta para el análisis la estructura de Python, para la generación de los códigos, que permiten realizar las siguientes pruebas:
 - **Juicio de expertos:** utilizada para la construcción de instrumentos de medición para asegurar que los ítems sean pertinentes y claros.
 - **Validez del contenido:** se refiere al grado en que los ítems de un cuestionario cubren todas las dimensiones del constructo en estudio.
 - **Índice de Validez de Contenido (IVC) de Lawshe.**
 - Estima los criterios de aceptación ($IVC \geq 0.58$ puede considerarse válido y se requiere 5 expertos).
 - Plantea el código para el IVC según la fórmula de Lawshe para la simulación del juicio de expertos.
 - Plantea el código para la Simulación de la Prueba Piloto del cuestionario a 50 estudiantes.
 - **Escala de Likert, en resumen:**
 - La escala Likert es ordinal, no intervalar.
 - Por tanto, no es válido matemáticamente calcular medias ni desviaciones estándar.
 - Solo se justifican calcular: mediana, moda o frecuencia.
 - **Matriz de correlaciones policóricas para el Análisis Factorial Exploratorio (AFE)**
 - El instrumento está medido con una escala de Likert de 5 puntos, la cual es ordinal.
 - Se utiliza la **matriz de correlaciones policóricas**, para ser riguroso en la medición, puesto que está diseñada para capturar la asociación entre variables ordinales bajo el supuesto de que hay una variable continua latente detrás de cada ítem.
 - **Análisis Factorial Exploratorio (AFE):**
 - Utiliza el Scree Plot, para encontrar los valores superiores a 1 para los primeros cuatro componentes principales.
 - Utiliza la rotación Varimax, para buscar cargas factoriales altas en un solo factor y bajas en los demás factores (4 factores) a fin de mejorar la interpretación de los datos.
 - **Evaluación de la Fiabilidad (Consistencia Interna)**
 - Realiza la medición de la fiabilidad y esto se refiere al grado en que un instrumento produce resultados consistentes bajo las mismas condiciones.
 - La consistencia interna en escalas tipo Likert es el Coeficiente Alfa de Cronbach (α) (que mide qué tan confiable es el instrumento). Este coeficiente estima la proporción de la varianza total de una escala que se debe a la varianza verdadera (no al error).
 - En resumen, un instrumento de levantamiento de información nuevo o en etapa de desarrollo, el Coeficiente Alfa de Cronbach (α) debe estar entre 0.6 y 0.7 podría ser aceptable. Pero para instrumentos ya establecidos se espera un valor mayor a α 0.8.
 - **Diseño Muestral Técnico:**
 - Explica que para el caso de estudio utilizó un diseño de muestreo probabilístico estratificado, el cual le permitió garantizar la representación de todos los dominios

- definidos. Y que el diseño contempla una afijación proporcional al tamaño del extracto es directamente proporcional al tamaño respecto del total de la población estudiantil objetivo.
- Realizó el cálculo de los factores de expansión, e indica que esto le permitió expandir los resultados de la muestra hacia la población de referencia, asegurando la validez de las estimaciones a nivel total y por estrato.
 - Consideraciones de la muestra, manifiesta que la selección de unidades dentro de cada estrato se asume la aleatoriedad simple. El diseño es unietápico y sin reemplazo e indica que no se consideran efectos de conglomeración.
- **Factor de expansión:**
 - Explica que un factor de expansión permite proyectar los resultados de la muestra hacia toda la población, convirtiendo los datos muestrales en estimaciones poblacionales.
 - El factor de expansión indica cuántas personas de la población están siendo representadas por una sola persona encuestada.
 - Cuando el valor del factor de expansión es alto, no es un error, pero sí indica que pocas personas están hablando por muchas personas.
 - Mientras más grande sea el factor, más cuidado se debe tener con los errores muestrales.
 - **Justificación metodológica para el uso de Análisis Factorial Exploratorio**
 - Explica que cuando el objetivo es construir un índice que represente un constructo latente como el bienestar universitario, el Análisis Factorial Exploratorio (AFE) es el método más adecuado, ya que permite modelar directamente las relaciones entre ítems y factores no observables, considerando además el error de medición asociado a cada ítem.
 - **Evaluación de Impacto con Regresión Discontinua - resultado**
 - Modelo Cuasi-experimental: Se presenta el modelo de regresión discontinua para medir el efecto causal de las becas.
 - Variables: Notas (variable de selección con punto de corte), Índice de Bienestar (variable de respuesta), Tratamiento (beca sí/no), y Variables de Control (campus, carrera, nivel).
 - Coeficiente de Impacto: Se enfatiza que solo el parámetro que mide el impacto es relevante para la toma de decisiones, simplificando el análisis tradicional econométrico.
 - Significancia de la Brecha: El objetivo es que la diferencia en el índice de bienestar entre los grupos con y sin beca, alrededor del punto de corte, sea estadísticamente significativa.

3. Preguntas y respuestas:

Aunque no se dedicó una sección específica de preguntas y respuestas al final de cada parte, las interacciones clave se integraron en el desarrollo de la sesión:

- **Preguntas a lo largo de la intervención:** por parte de la audiencia se realizaron varias preguntas de parte de los participantes, las mismas que fueron solventadas por el instructor de forma clara y precisa. Además, el instructor dio apertura a que, si tienen consultas específicas de trabajos que estén realizando los asistentes, le pueden escribir a su correo para solventarlas, dando apertura al aprendizaje de forma práctica y específica.
- **Interés del público en la regresión discontinua:** A pesar de las limitaciones de tiempo, el público solicitó una explicación detallada del modelo de evaluación de impacto, lo que llevó al instructor Andrés Galvis a extender la sesión para abordar este tema. Esto demuestra el alto nivel de interés y la relevancia del contenido para la audiencia.
- **Aportes de la audiencia:** Se agradeció la participación de quienes aportaron correcciones y confirmaron la ejecución de los códigos, generando un ambiente de aprendizaje interactivo.

4. Conclusiones y Próximos Pasos

El evento concluyó con varias ideas fuerza y la proyección de futuras actividades:

- Se destacó la importancia de usar metodologías estadísticas y de ciencia de datos rigurosas para evaluar el impacto de las intervenciones públicas, garantizando la asignación eficiente y justa de los recursos.
- La construcción de un índice requiere múltiples etapas de validación (contenido, criterio, estadística, fiabilidad), no solo una, para asegurar la calidad del instrumento.
- Se vislumbra un futuro donde la inteligencia artificial, especialmente a través de modelos multi-agente, podría optimizar procesos como el juicio de expertos, dado su vasto volumen de información.
- Una llamada a las instituciones clave (Secretaría de Planificación, INE) para que establezcan directrices claras y homogéneas para la construcción de indicadores a nivel nacional.

5. Observaciones

- El formato del taller fue muy efectivo al combinar la explicación teórica de la construcción de índices y evaluación de impacto con la implementación prácticas. Esto permitió a la audiencia no solo entender los conceptos, sino también visualizar su aplicación en un caso real.
- La discusión sobre la medición de las escalas Likert fue un punto y muy didáctico, destacando la importancia de la fundamentación matemática para su medición. En la que se deja claro que no se puede calcular medias ni desviaciones estándar y solo se justifica calcular mediana, moda o frecuencia. Y se deja claro que es de escala es ordinal y no intervalar.
- Se deja en claro que evaluar no es lo mismo que monitorear. Monitoreo/Seguimiento es como llevar un registro y muestra lo que planifico y lo que ocurrió. En cambio, evaluar va más allá, se pregunta “¿qué hubiese pasado si no hubiéramos intervenido?” o ¿Realmente cambio algo gracias a la

intervención del estado? Y para responder, se necesitan métodos estadísticos para saber si la política/intervención han funcionado o no.

- El coeficiente de impacto es el único relevante en la regresión discontinua es un mensaje poderoso para los tomadores de decisiones, simplificando la interpretación y la acción.
- El expositor vinculó constantemente los conceptos estadísticos con la necesidad de justificar la asignación de recursos y la rendición de cuentas ante organismos como la Contraloría, lo que resalta la aplicabilidad y el valor de estos conocimientos en el sector público.
- Al compartir el código del caso de estudio y los materiales, se facilita que los asistentes puedan replicar el análisis, experimentar y aprender por sí mismos, lo que es un pilar fundamental en las capacitaciones al aprender haciendo.

Este informe destaca la riqueza del contenido presentado y la forma efectiva en que se abordaron temas complejos de estadística y ciencia de datos, poniéndolos al servicio de la evaluación de políticas públicas para una gestión más eficiente.