**Propuesta para Temario de Entrenamiento en Usos Estadísticos usando los Datos del Barómetro de las Américas**

Esta propuesta tiene varios objetivos. En primer lugar, se propone *difundir* el proyecto del Barómetro de las Américas, mediante el uso de los datos por diversos públicos. También se propone *educar* a comunidades de impacto en los usos estadísticos de los datos del Batómetro de las Américas para reportes e investigaciones. Finalmente, se propone *reproducir* resultados de informes y reportes producidos por LAPOP y así promover la transparencia de la investigación.

LAPOP usa comúnmente el software STATA para sus análisis y reportes. En este temario se propone el uso del software R y de la plataforma RStudio, aunque la propuesta está abierta a trabajar en ambos lenguajes.

***Estado de la oferta de entrenamiento***

En la web se pueden encontrar innumerables recursos sobre el lenguaje R y sus aplicaciones en las áreas de estadística y ciencia de datos. Por ejemplo, las plataformas Coursera[[1]](#footnote-1) y edX[[2]](#footnote-2) cuentan con programas especializados en el uso de R para estadística y ciencia de datos. Estos cursos ofrecen un entrenamiento general en el uso de R para múltiples propósitos, en diferentes campos de especialización.

En ciencias sociales, y en ciencia política en particular, hay también una oferta de entrenamiento. Por ejemplo, la European Consortium for Political Research (ECPR) ofrece un curso de R básico como parte de su Escuela de Verano[[3]](#footnote-3).

También existe una oferta no estructurada en repositorios web. Por ejemplo, la página *Quantitative Politics with R* ofrece un temario que va desde lo básico, como instalación del software R y RStudio, pasando por el manejo de datos y la visualización, hasta el modelamiento y la presentación de resultados[[4]](#footnote-4). Los ejemplos usan datos agregados electorales de Estados Unidos, la European Social Survey, la General Social Survey, la American National Election Study, el Comparative Manifesto Project y los datos del proyecto Varieties of Democracy.

De otro lado, se tienen publicaciones dedicadas al entrenamiento en R en el campo del análisis político. Recientemente, por ejemplo, se publicó *R for Political Data Analysis*[[5]](#footnote-5). Una propuesta bastante completa viene del libro *Quantitative Social Sciencie: An Introduction* (Imai 2017). Esta publicación presenta una variedad de temas estadísticos, que van desde la descripción y el modelamiento hasta el análisis de redes y el análisis espacial. Esta propuesta es interesante debido a que viene acompañada de un repositorio en GitHub, desde donde se puede descargar las bases de datos que se usan a lo largo de los temas[[6]](#footnote-6). Un aspecto resaltante es que el autor hace uso de bases de datos referentes a artículos académicos publicados en ciencia política. De esta manera, esta publicación sirve para que los estudiantes reproduzcan resultados de trabajos publicados.

***Propuesta***

Los aspectos diferenciadores de esta propuesta son:

* Se trabajaría solo con datos del Barómetro de las Américas de LAPOP. Esto le daría unidad y coherencia a todo el temario. La mayor parte de la oferta de entrenamiento usa diversas fuentes de información por cada tema.
* No existe en la actualidad algún proyecto de opinión pública que haya desarrollado una plataforma educativa para el uso de sus propios datos[[7]](#footnote-7).
* Este temario sería un complemento del Data Playground. Algunos resultados presentados en los reportes pueden ser reproducidos en el Data Playground y, a su vez, se puede instruir en cómo calculados en R[[8]](#footnote-8).
* Mejoraría la diversidad de recursos de libre acceso disponibles en la página web de LAPOP.
* Se trabajaría con datos de la realidad latinoamericana, público al que va dirigido esta propuesta. La mayor parte de la oferta de entrenamiento se basa en datos de la realidad norteamericana y europea.
* Los materiales de entrenamiento estarían en español[[9]](#footnote-9). Si bien existe una oferta de entrenamiento estadístico en español, esta no es estructurada y no es particular a las ciencias sociales o a la opinión pública o al campo de la cultura política.
* Se usarían reportes y producción propia de LAPOP para los ejemplos del temario. Esto brindaría mayor difusión a estos documentos y fomentaría la transparencia en la investigación. Los primeros materiales han empleado ejemplos del reporte “El Barómetro de las Américas de LAPOP toma El Pulso de la Democracia” y de la serie *Insights*.
* Se usaría el software de libre acceso R, lo que permitiría una mayor difusión del análisis de datos entre comunidades de impacto a las que se le dificulta el acceso a software propietario.

***Proceso***

El desarrollo del temario implica los siguientes pasos:

1. Elaboración de un documento en RMarkdown[[10]](#footnote-10) para cada tema. Este documento se puede actualizar y modificar continuamente.
2. Cada documento en RMarkdown permite la importación a formatos html y a pdf.
3. Todos estos materiales se suben a un repositorio en GitHub (en la cuenta de LAPOP-Central). Las versiones preliminares se pueden almacenar en un repositorio provisional y las versiones finales en el repositorio para distribuir.
4. El repositorio para distribuir:
   1. Está ordenado por temas: en cada carpeta se encuentra el documento base RMarkdown, el archivo producido en html y en pdf.
   2. Incluye una versión reducida de la base de datos del *merge* del Barómetro de las Américas de la ronda 2018.

***Distribución de materiales***

Con estos documentos producidos se puede:

* Estos materiales pueden ser la base para la creación de un website que explique el objetivo de estos materiales y guie a los interesados en los temas. Un ejemplo de esta website se puede ver provisionalmente en el siguiente enlace: <https://arturomaldonado.github.io/BarometroEdu_Web/index.html>.
* Compartir el link al repositorio de GitHub (<https://github.com/lapop-central/materials_edu.git>).
  + Personas con cuenta en GitHub pueden “clonar” este repositorio en sus cuentas personales.
  + Personas sin cuenta en GitHub puedes descargar los documentos del repositorio y leerlos en R.
* Estos materiales también se pueden subir a un repositorio de documentos de R, llamado RPubs. Esos documentos están disponibles provisionalmente es: <https://rpubs.com/arturo_maldonado>.

***Temario propuesto***

Esta propuesta plantea desarrollar un temario inicial sobre estadística básica, el que, paso a paso, pueda ir complejizándose. Cada uno de estos temas se desarrollarán en una serie de documentos, los que serán de libre acceso, y que pueden estar en formato pdf o html y que puedan ser subidos fácilmente en la página web de LAPOP.

Como un avance se ha desarrollado una serie de documentos que utilizan los datos del Barómetro de las Américas para reproducir ejemplos de estadística descriptiva e inferencial. Para estos ejemplos se uso el último reporte regional “El Pulso de la democracia” y, en particular, la sección sobre redes sociales y actitudes políticas.

Se propone el siguiente temario de entrenamiento:

| Título | Temas generales | Tema específico | Fuente |
| --- | --- | --- | --- |
| Manejo de datos usando los datos del Barómetro de las Américas por LAPOP | Recodificación  Calcular una variable  Seleccionar casos | Recodificación de variables a escala 0-100  Cálculo del índice de tolerancia política  Seleccionar países para análisis | Reporte “El Pulso de la Democracia” |
| Análisis con datos ponderados usando los datos del Barómetro de las Américas por LAPOP | Uso de la librería “survey” para ponderaciones | Inclusión de variable “weight1500” en los cálculos descriptivos e inferenciales |  |
| Estadística descriptiva con datos del Barómetro de las Américas por LAPOP (1) | Descripción de datos categóricos nominales  Gráficos univariados | Descripción y gráficos de uso de redes sociales: Facebook, Twitter y Whatsapp | Redes sociales y actitudes políticas del reporte “El Pulso de la Democracia” |
| Estadística descriptiva con datos del Barómetro de las Américas por LAPOP (2) | Descripción de datos categóricos ordinales  Tablas cruzadas | Frecuencia de uso de redes sociales: Facebook, Twitter y Whatsapp  Uso de redes sociales por país | Redes sociales y actitudes políticas del reporte “El Pulso de la Democracia” |
| Estadística descriptiva con datos del Barómetro de las Américas por LAPOP (3) | Descripción de datos numéricos | Promedios y dispersión de edad y educación de usuarios de redes sociales | Redes sociales y actitudes políticas del reporte “El Pulso de la Democracia” |
| Intervalos de confianza con datos del Barómetro de las Américas por LAPOP | Construcción de intervalos de confianza totales y por grupos.  Gráficos | Satisfacción con la democracia por grupos de consumo de información política en redes sociales | Redes sociales y actitudes políticas del reporte “El Pulso de la Democracia” |
| Pruebas de significancia usando los datos del Barómetro de las Américas por LAPOP (1) | Prueba t de diferencia de 2 medias |  |  |
| Pruebas de significancia usando los datos del Barómetro de las Américas por LAPOP (2) | ANOVA |  |  |
| Asociaciones categóricas usando los datos del Barómetro de las Américas por LAPOP | Prueba de chi cuadrado |  |  |
| Correlación usando datos del Barómetro de las Américas por LAPOP | Diagrama de dispersión  Análisis de correlación | Correlación entre medida de apoyo a la democracia en LAPOP y índice V-Dem. | Redes sociales y actitudes políticas del reporte “El Pulso de la Democracia” |
| Regresión usando los datos del Barómetro de las Américas por LAPOP (1) | Diagrama de dispersión  Modelo de regresión lineal simple  Presentación de resultados | Correlación y modelamiento de la relación entre la insatisfacción con los servicios de salud, la evaluación del presidente y la tasa de letalidad por COVID-19 |  |
| Regresión usando los datos del Barómetro de las Américas por LAPOP (2) | Modelos de regresión lineal múltiple  Presentación de resultados |  |  |
| Modelos lineales generalizados usando los datos del Barómetro de las Américas por LAPOP | Modelos de regresión logística.  Presentación de resultados |  |  |
| Inferencia causal y análisis de experimentos usando datos del Barómetro de las Américas por LAPOP |  |  |  |
| Reporte de resultados usando los datos del Barómetro de las Américas por LAPOP | Uso de RMarkdown para generar reportes | Producción de documentos de reporte de resultados en RMarkdown para exportar a pdf o a html. |  |
| Mapas y gráficos espaciales usando los datos del Barómetro de las Américas por LAPOP | Uso la librería “leaflet” para presentar mapas de resultados | Mapas regionales con promedios de variables del Barómetro de las Américas  Uso del Data Playground |  |
| Gráficos interactivos usando los datos del Barómetro de las Américas por LAPOP | Uso de la librería “plotly” |  |  |
| Creación de dashboards y apps usando los datos del Barómetro de las Américas por LAPOP | Uso de la librería “shiny” |  |  |

***Documentos disponibles***

|  |  |
| --- | --- |
| Tema | Disponible en: |
| Manejo de datos usando los datos del Barómetro de las Américas por LAPOP | * <https://arturomaldonado.github.io/BarometroEdu_Web/Manipulacion.html> * <https://rpubs.com/arturo_maldonado/727887> * <https://github.com/lapop-central/materials_edu.git> |
| Análisis de datos usando factores de expansión | * <https://arturomaldonado.github.io/BarometroEdu_Web/Expansion.html> * <https://rpubs.com/arturo_maldonado/728626> * <https://github.com/lapop-central/materials_edu.git> |
| Estadística descriptiva usando variables nominales | * <https://arturomaldonado.github.io/BarometroEdu_Web/Descriptivos.html> * <https://rpubs.com/arturo_maldonado/696770> * <https://github.com/lapop-central/materials_edu.git> |

***Resultados***

* Mayor número de descargas de las bases de datos.
* Mayor uso de los datos del Barómetro de las Américas.
* Mayor número de descargas de los reportes y artículos que usan los datos del Barómetro de las Américas.
* Mayor difusión de los productos de LAPOP entre comunidades de impacto, como estudiantes y jóvenes investigadores.
* Mayor reconocimiento de LAPOP y del Barómetro de las Américas como la encuesta de cultura política más importante de la región.

***Posibilidades futuras***

* Estos documentos pueden servir como una base para ofrecer talleres de capacitación a universidades públicas y privadas en América Latina.
* Estos documentos pueden servir como un guion para el desarrollo de materiales audiovisuales.

1. Ver: <https://es.coursera.org/specializations/statistics>. [↑](#footnote-ref-1)
2. Ver: <https://www.edx.org/es/course/statistics-and-r>. [↑](#footnote-ref-2)
3. Ver: <https://ecpr.eu/Events/Event/PanelDetails/8411>. [↑](#footnote-ref-3)
4. Ver: <http://qpolr.com/>. [↑](#footnote-ref-4)
5. Ver: <https://www.routledge.com/R-for-Political-Data-Science-A-Practical-Guide/Urdinez-Cruz/p/book/9780367818890>. [↑](#footnote-ref-5)
6. Ver: <https://kosukeimai.github.io/qss-package/index.html>. [↑](#footnote-ref-6)
7. Una excepción puede ser el Pew Research Center, que ha desarrollado un paquete propio en R y en Python, especializado en el análisis de datos de encuestas. Este paquete permite realizar recodificaciones, mostrar estimados ponderados de variables categóricas, entre otras funciones. Este centro de investigación ha descrito algunas de estas funcionalidades en entradas de blog (ver: <https://medium.com/pew-research-center-decoded/analyzing-international-survey-data-with-the-pewmethods-r-package-3b0b21cba607>). Sin embargo, no se trata de un temario estructurado, pero comparte con esta propuesta el espíritu de informar a los lectores cómo analizar los datos producidos por este centro usando software libre. [↑](#footnote-ref-7)
8. Una primera prueba de la reproducción de resultados mediante el Data Playground y el software R se dio en la II Conferencia Anual “El Público y la Democracia en las Américas”, donde se ofreció un taller en el uso de los datos del Barómetro de las Américas. [↑](#footnote-ref-8)
9. Estos materiales pueden ser traducidos al inglés para ampliar la difusión. [↑](#footnote-ref-9)
10. Rmarkdown no es el documento nativo de R para hacer scripts. Este tipo de documento es más cercano a Latex, pues permite introducir código de R y mostrar, en el mismo documento, el producto del código: datos, tablas, gráficos, etc. Además, no solo permite producir documentos, sino también presentaciones, dashboards y webpages. Ver: <https://rmarkdown.rstudio.com/>. [↑](#footnote-ref-10)