

DESCRIPCIÓN DEL PROYECTO

La Secretaría de Salud ha decidido contratar a una consultoría de análisis estadístico, en la cual usted trabaja, para conocer el fenómeno relativo al **número de casos por Diabetes *melitus* tipo II (no insulino-dependiente)** con el objeto de tomar las medidas pertinentes a nivel nacional.

En particular, le interesa conocer el fenómeno a partir de la información que ha estado generando desde el 2003 hasta el 2013 al siguiente nivel de desglose:

- Grupo de edad.
- Tipo de fuente.
- Mes.

La información puede descargarse de la siguiente página:

<http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>.

Objetivo

El propósito de este proyecto es la de trabajar en un proyecto real desde la extracción de los datos hasta la presentación final de los resultados utilizando las herramientas y conceptos vistos en clase.

En esta primera etapa sólo se pretende conocer la información desde el punto de vista exploratorio con el objeto de describir la información y organizarla para la etapa inferencial.

Estructura y consideraciones

El proyecto deberá cumplir con los siguientes puntos:

1. Tendrá que ser **totalmente reproducible**, salvo la generación de los archivos de texto plano (aunque deberá documentarse en el archivo README.md y describirse en el reporte técnico) y los reportes con el análisis.
2. Considerando que la información bruta se encuentra en archivos con formato PDF, deberán capturarse o convertirse en archivos de texto plano (CSV o TXT). Recuerde que son 99 archivos considerando los tres niveles de información, los años de observación y el sexo (incluyendo la población general).

Nota: En la práctica, por lo general, la información no se encuentra en el formato adecuado, se sugiere utilizar la solución descrita en el artículo: “[How to Make Government Data Sites Better](#)”.

3. Deberá entregarse una carpeta bajo control de versiones (i.e., deberá existir el folder .git con al menos dos commits), con base en la organización descrita en la [sesión 08](#).

Nota: No se requiere “subirla” a GitHub. Sin embargo, al ser un trabajo en equipo, se recomienda usar una cuenta privada en <https://bitbucket.org/>.

4. Se entregarán dos documentos en formato Word o PDF: Un **reporte técnico** y el **resumen ejecutivo**, este último no deberá exceder de cinco páginas y tendrá que basarse en la información más relevante del primero dirigido a una audiencia que posee un conocimiento medianamente técnico (e.g., sabe qué es una desviación estándar y puede interpretar un diagrama de caja y brazos).

En el reporte ejecutivo, considere lo aprendido en la lectura “Statistical Narrative - Telling Compelling Stories with Numbers” de Stephen Few.

Por otra parte, el documento técnico deberá contener al menos la siguiente información:

- **Introducción:** El objeto de esta sección es la de proveer una introducción al conjunto de datos, contexto, objetivo y contenido del resto del documento. No deberá ser mayor a dos páginas.
 - **Definición del problema:** Deberán definirse las preguntas que pretenden responderse con el análisis exploratorio de los datos, indicando la razón por la cuál se eligieron y resultan relevantes para el cliente (i.e., la Secretaría de Salud).
 - **Obtención y descripción de los datos:** Esta sección deberá incluir una descripción de la base “ideal” a fin de responder las preguntas establecidas en la definición del problema y las fuentes de información disponibles a su alcance (además de la descrita en el contexto). Con base en lo anterior, deberá establecer los alcances de los datos y sus limitantes. También deberá incluir una descripción de las variables indicando su tipo y escala.
 - **Calidad de los datos:** Deberá incluir un análisis sobre la calidad de los datos considerando las pruebas pertinentes. La prueba de cifras de control tendrá que realizarse (mediante código) para cada año y nivel de información (i.e., General = Femenino + Masculino).
 - **Transformación de los datos:** Tiene como objeto describir el procesamiento y recodificación de las variables a fin de obtener los datos procesados.
 - **Análisis exploratorio:** En esta sección se incluirán las gráficas, tablas y medidas descriptivas utilizadas para responder las preguntas de interés.
 - **Conclusiones:** Finalmente, se incluirá un resumen de los hallazgos. En vez de repetir los resultados descritos en las secciones anteriores, trate de “hilar” los resultados en una historia consistente. Cuestione los resultados obtenidos y proponga ideas para un análisis posterior.
5. Deberá incluirse, en la carpeta del proyecto, un archivo en formato .RData con los datos procesados (mismos que deberán cumplir con las propiedades de tablas organizadas, ver [sesión 10](#)), así como un diccionario de datos.
6. El código deberá documentarse considerando las recomendaciones vistas en clase.

Entrega

- El folder del proyecto **deberá enviarse al correo descrito en el temario** (que es el mismo que se encuentra en el pie de página de este documento), mientras que **el reporte y el resumen ejecutivo deberán entregarse impresos**.
- La **fecha de entrega** del proyecto será el **miércoles 25 de marzo**.
- El proyecto deberá entregarse en equipo de **máximo tres personas**.