

# EST729: Tarea #2

Christian Araya  
christian.araya@pucv.cl

PUCV — September 1, 2020

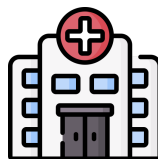
## 1 Introducción

La base de datos corresponde al cruce de la siguiente información: los resultados de la aplicación de la encuesta adjunta en Anexos y la asociación del ID del paciente con el monto pagado por la transacción que gatilló su participación en la encuesta de pacientes ambulatorios. El valor corresponde al monto privado, es decir, antes de aplicar bonificación de acuerdo a la previsión del paciente.

Analice cuidadosamente la base de datos y note que algunas variables fueron recodificadas con la intención de facilitar el análisis estadístico. Además, notará que existe información perdida.

## 2 Actividades

1. (10%) Describa la muestra realizando Análisis Exploratorio de Datos. Seleccione las características más importantes que usted desee mostrar. **Importante:** no necesita realizar un análisis descriptivo de las 21 variables en escala Likert para contestar esta pregunta.
2. (20%) Enfóquese en las preguntas 5, 8, 14 y 18, medidas en Escala Likert, y en el Monto de la Prestación. Diagnostíquelas en términos de la cantidad de datos perdidos y su relación con el resto de variables al momento de ocurrir la pérdida de información. En este punto, considere que las preguntas tipo Likert arrojan resultados en escala Ordinal, mientras que Monto de la Prestación en escala de razón. **Esto limitará los análisis que usted pueda realizar.**
3. (40%) Impute los datos perdidos (en las 5 variables antes destacadas) empleando: 1) promedio por variable; 2) algún refinamiento del promedio que estime apropiado; 3) mediante imputación aleatoria (ver **Anexo**). Analice a través de alguna medida de desempeño el impacto de la técnica empleada y compare. **Importante:** en este punto, no pierda de vista que los resultados de las preguntas tipo Likert se obtienen en escala Ordinal.
4. (15%) Estudie cuál es el impacto de la técnica de imputación escogida si usted calculara el grado de asociación entre las preguntas de la encuesta. Si usa Pearson, estará aproximando los resultados. La opción correcta es aplicar la correlación de Spearman cuando existen variables Ordinales.
5. (15%) Impute los datos perdidos en Monto de la Prestación a través de una regresión lineal con alguna o algunas preguntas tipo Likert que usted considere pertinentes. Diagnostique el impacto de la imputación realizada a través de algún indicador clave. ¿Desde un punto de vista general, cree usted que tiene sentido la aplicación de esta técnica?



### 3 Imputación aleatoria simple

Un enfoque común y lógico para realizar imputación consiste en generar un valor aleatorio basándose en la información existente para la variable. El proceso se asemeja a realizar un muestreo con reemplazo a partir del set de datos disponibles.

Para ello, en el desarrollo de esta tarea, plantee en R una función que le permita generar un valor aleatorio para imputar un valor perdido: 1) si la variable a imputar corresponde a una respuesta a una pregunta tipo Likert, considere las probabilidades que determinan la información observada; 2) en el caso del Monto de la Prestación, puede recurrir a una distribución continua en caso de que los datos existentes respondan a alguna distribución que usted conozca.