

EST711: Tarea #2

Christian Araya
christian.araya@pucv.cl

PUCV — August 17, 2021

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 copago, es decir, después 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.
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 alguna técnica de tipo *hot deck*; 4) a través de regresión lineal, considerando las 3 últimas preguntas de la encuesta. Analice el impacto de las técnicas empleadas y compare, focalizando el análisis en medidas de resumen (univariadas).
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. **Nota:** si usa Pearson, estará aproximando los resultados. La opción correcta es aplicar la correlación de Spearman cuando existen variables Ordinales.
5. (15%) Investigue y utilice alguna técnica de imputación no cubierta en la pregunta 3) y dimensione el efecto de esta técnica. ¿Qué concluye a partir de los resultados obtenidos?