
Trabajo Práctico 1

EcoBici

Un estudio sobre el uso de bicicletas públicas en la Ciudad de Buenos Aires

Métodos Estadísticos Aplicados a los Negocios

Fecha de entrega: 19 de abril de 2020

1. INSTRUCCIONES

El trabajo es grupal (con un máximo de 4 alumnos). Deberán entregar un informe final **a través del campus virtual** en el cual adjuntarán el Script en R y las bases utilizadas para poder replicar sus resultados. Entreguen un único ejemplar por grupo. Tienen tiempo para entregar el TP hasta el **domingo 19 de abril a las 23.55hs**. No se aceptan entregas tardías. En dicho caso, el trabajo estará desaprobado.

2. CONSIGNA

La creación de las bicisendas en la Ciudad de Buenos Aires empezó en 1996 a partir del reclamo de distintas asociaciones de ciclistas que requerían sendas exclusivas para poder circular. A principios de 1997, la Legislatura porteña llamó a licitación para construir las ciclovías y en septiembre de 1998, se inauguró el primer tramo de casi 8 kilómetros entre los barrios de Palermo y Belgrano.

Actualmente, la Ciudad de Buenos Aires tiene su Sistema de Transporte Público de Bicicletas con estaciones automáticas que ofrece en forma gratuita, las 24 horas, todos los días del año. La red de ciclovías cubre aproximadamente 250km y el sistema de EcoBici cuenta con cerca de 400 estaciones.

Este servicio se ofrece como una forma alternativa de transporte dentro de la Ciudad que resulta rápido, ecológico y saludable.

En este trabajo práctico usted trabajará con los datos públicos de usuarios y recorridos que se publican en <https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>.

Cada pregunta vale 10 puntos. Se pide:

- (1) Descargue los datos de 2018 y a través de gráficos/tabulaciones describa brevemente las principales estadísticas relacionadas al uso de la EcoBici.
- (2) Para dar un mejor servicio al usuario, se desea estudiar si (estadísticamente) existe estacionalidad en el uso de la EcoBicis. Una forma de probarlo (y comunicarlo) es construir un gráfico que reporte los intervalos de confianza de la cantidad media de usuarios por cada día de la semana. ¿Hay evidencia estadística de estacionalidad en el uso? Comente.
- (3) Es sabido que la media se puede ver afectada por la presencia de valores extremos (e.g. feriados). Sobre los datos anteriores se pide que descuenta del análisis los días feriados de 2018 y vuelva a construir el gráfico. ¿Cambian los resultados? Comente.
- (4) Concéntrese en el usuario 606320. en 2018 realizó 95 recorridos. Elimine los casos en que la estación de origen coincide con la de destino para el mismo recorrido. Queremos calcular la velocidad media a la que usualmente se desplaza este ciclista. Cuando trabajamos con velocidades lo correcto es utilizar la media armónica (y no la media aritmética). Calculando la distancia en bici entre las estaciones (use Google Maps) y con el dato de la duración del recorrido (asumiendo que la persona “no interrumpe el recorrido”, lo cual para este usuario resulta razonable), calcule la velocidad media de circulación de dicho usuario. Comente las ventajas de usar la media armónica en lugar de la media aritmética.
- (5) Una de las estaciones más concurridas es la “009-PARQUE LAS HERAS” y desde allí la mayoría de los usuarios suele dirigirse hasta la estación “066-BILLINGURST”. Según Google Maps, entre una estación y otra hay 1.5km y en promedio en bici debería tardarse 8 minutos. Se considera que si un usuario tarda más de 15 minutos es porque hizo escalas en el camino u optó por otro recorrido antes de dejar la bici en su destino final. Evalúe empíricamente si la proporción de usuarios que tardan más de 15 minutos no supera el 20%.
- (6) Sobre la prueba del punto anterior obtenga la curva de potencia del test. Comente.
- (7) En mayo de 2018 se inauguró la estación Facultad de Derecho de la línea H de subtes. ¿Afectó esto al uso de la EcoBici por la zona? Evalúe empíricamente si en la estación “001-FACULTAD DE DERECHO” se registró una diferencia estadísticamente signifi-

cativa en el uso de las EcoBicis antes y después de la inauguración de la estación de subte. Comente.

- (8) Probar si existen diferencias estadísticamente significativas en el uso del sistema de EcoBici por género.
- (9) Probar lo mismo que en el punto anterior, pero por grupo etario.
- (10) **¡Vamos por el 10! Bootstrapping:** dado que la distribución del coeficiente de correlación no es normal ya que está limitada entre -1 y 1, el método bootstrap puede ser muy útil para hacer inferencia sobre dicho parámetro. Para ello, nos centraremos en la correlación entre la edad del usuario de EcoBici y la duración del recorrido para aquellos individuos que hacen el recorrido “Parque Las Heras - Billingurst” (mismos que en el punto 5 del TP). Siga los siguientes pasos para construir el intervalo de confianza del coeficiente de correlación entre dichas variables:
 - Defina una función (`function()`) que permita calcular el estadístico sobre el cual se quiere realizar el bootstrap.
 - Utilice la función `boot()` usando la base de datos de recorridos y la función generada en el paso anterior para tomar 1.000 muestras con reemplazo.
 - Obtenga el histograma de las estimaciones del bootstrap. Describa.
 - Utilice la función `boot.ci()` con la opción "perc" y obtenga el intervalo de confianza al 95%. Interprete. ¿Resulta estadísticamente significativa la correlación entre la duración del recorrido y la edad del ciclista?