

Parcial computacional sobre Test de Hipótesis Patentes

Métodos Estadísticos en Física Experimental

2023

1 Enunciado

Desde el 1 de abril de 2016, los automóviles de nuestro país se patentan con un nuevo sistema, este es de la forma $L_1L_2 N_1N_2N_3 L_3L_4$ donde L son letras y N son números. Así, esta nueva notación, que comenzó con AA000AA, luego AA000AB, ..., AA000ZZ, AA001AA, etc, terminará con ZZ999ZZ.

Hay muchas preguntas que uno podría pretender responder con sólo mirar patentes, y lo que aprendimos en MEFÉ nos puede ayudar a hacerlo. Para eso les pedimos que observen en la calle y anoten algunas patentes que comiencen con AF o AG. Lo primero que deberás hacer es transformar cada una de las k patentes observadas en un número natural, asignando el 1 a la patente AF000AA, el 2 a la AF000AB y así sucesivamente, hasta llegar a m , el natural correspondiente a la patente más nueva que forme parte de la muestra. Una vez que tengas la lista, estarás en condiciones de empezar el parcial.

Dato: Podés usar que la patente más nueva en circulación cuando se tomó la muestra era AG084DK.

2 ¿Uniformemente distribuidos?

Tenemos razones para pensar que las patentes observadas (o mejor dicho, los naturales que les asociaste) representan una variable aleatoria con distribución uniforme.

1. Aplicá el test de Kolmogorov-Smirnov sobre la muestra (aunque vale para distribuciones continuas trabajaremos en esa aproximación) y presentá en una misma figura: la distribución teórica y la experimental, el estadístico observado, el p-valor, y decinos si con una significancia $\alpha = 0.05$ se puede rechazar la hipótesis nula. Sugerencia: No uses un test ya implementado en tu lenguaje de programación, codealo vos. De hecho, en el ítem siguiente no vas a poder evitar tener que hacerlo.
2. Considerá ahora, como hipótesis alternativa, que la distribución de patentes es exponencial de parámetro $\lambda=4 \times 10^{-7}$ y calculá la potencia del test del ítem anterior. Ayuda: No olvides que la potencia es una propiedad del Test y no de los datos. Y tampoco olvides que el test queda definido por su H_0 y su significancia. Entonces, si H_1 es verdadera, los datos provendrían de una exponencial. Sin embargo al aplicar el test, la distribución acumulada experimental sería igualmente comparada con

la acumulada de la uniforme (de otro modo hubiera cambiado el test!). Averiguá entonces la distribución del estadístico cuando H_1 es verdadera haciendo una simulación y luego, usando el valor crítico del estadístico hallado en el ítem anterior (para $\alpha=0.05$) podrás calcular la potencia del test.

3 ¿Independiente del barrio?

En este punto debes tomar dos grupos de patentes correspondientes a barrios diferentes. Es importante que elijas dos barrios bien alejados.

1. Aplicá el test de Wilcoxon para testear si ambas muestras provienen de poblaciones con la misma esperanza y calculá el p-valor (p_w). Para más información sobre este test ver el libro de la señora Frodesen página 450.
2. Ahora usá el estadístico propuesto en el problema 4 de la guía de Test de Hipótesis. Pero antes, notá que no es correcto usar que tiene distribución t-student. ¿Por qué? Tomando como hipótesis nula que la distribución de patentes es uniforme, encontrá computacionalmente la distribución del estadístico propuesto.
Ayuda: La idea es que generes N veces dos set de datos con distribución uniforme y que calcules el valor del estadístico U . Así tendrás N valores para U y conocerás su distribución. Luego podrás aplicarle el test a tus datos y usar esa distribución para calcular el p-valor.
3. Aplicá el test construido en el ítem anterior (ahora que si conoces la distribución de su estadístico) sobre tus observaciones y calculá el p-valor (p_t).

4 Generalidades

1. Indicar nombre, apellido, libreta o DNI en el informe.
2. Dar una descripción clara y precisa de la metodología utilizada.
3. Incluir todos los gráficos como figuras con sus correspondientes leyendas.
4. Justifique las hipótesis en la que se sustenten sus resultados y discuta los resultados obtenidos.
5. La fecha límite para la entrega es el martes 11 de julio a las 11 hs.
6. Enviar por mail a dariorodriguesfm@gmail.com. Nombrar al archivo de la siguiente manera: TH-SuApellido.pdf y adjuntar en el mismo mail los códigos desarrollados. Utilizar como asunto del mail: "TH-Patentes".
7. Por consultas sobre interpretación de los enunciados escribir al grupo: mefe-1c-2023@googlegroups.com, así todos tienen acceso a todas las respuestas.