

PROBABILIDAD y ESTADÍSTICA (61.09 - 81.04)

Fecha de entrega: 10 de Diciembre 23:59hs

2do cuatrimestre – 2023

Apellido y Nombres:

Padrón:

En el mercado de smartphones, uno podría pensar que los dispositivos con mayor capacidad de almacenamiento suelen tener baterías más duraderas, dado que suelen ser modelos más premium con mejores especificaciones, incluida la batería. Modelar estos datos podría ayudar a estimar la duración de la batería en función de su capacidad de almacenamiento, algo útil para los consumidores al elegir un nuevo dispositivo. El archivo `smart.txt` posee valores registrados sobre capacidad de almacenamiento (primera columna en GB) y la respectiva duración de baterías (segunda columna en horas). Utilizando Python o R resolver:

1. Antes de suponer una distribución conocida para cada variable, estimar las varianzas de forma insesgada (y por separado).
2. Asumiendo que la distribución de la capacidad de almacenamiento es normal, se desea hacer un test para rechazar que la media es μ_0^x . Graficar el p-valor en función de μ_0^x . Relacionar dicho gráfico con el concepto de nivel de significación. ¿Qué puede decir del punto donde el p-valor alcanza el máximo?
3. Graficar la función de distribución empírica de la capacidad de almacenamiento y compararla con la curva correspondiente a una normal cuya media corresponda al valor que maximiza el p-valor del inciso 2 y cuya varianza sea la estimada en el inciso 1.
4. Antes de asumir una distribución conocida para la duración de las baterías, se desea hacer un test para rechazar que la media es μ_0^y . Graficar el p-valor asintótico en función de μ_0^y . Relacionar dicho gráfico con el concepto de nivel de significación asintótico. ¿Qué puede decir del punto donde el p-valor asintótico alcanza el máximo?
5. Graficar el histograma de la duración de las baterías y compararla con la curva correspondiente a una normal cuya media corresponda al valor que maximiza el p-valor asintótico del inciso 4 y cuya varianza sea la estimada en el inciso 1.
6. Asumiendo que los datos corresponden a una normal bivariada cuyas medias y varianzas son las utilizadas en los incisos 3 y 5 (asumiendo que son los verdaderos valores y son conocidos), graficar la log-verosimilitud en función de ρ . Estimar por máxima verosimilitud el coeficiente de correlación.



Recomendamos seguir los siguientes pasos:

- Halle una expresión analítica (a mano) para la log-verosimilitud $\log(L(\rho)) = \sum_{i=1}^n \log f_{\rho}(x_i, y_i)$.
 - Defina una función que para cada ρ devuelva el valor de $\log(L(\rho))$.
 - Construir el gráfico pedido. Para evitar errores numéricos suponer $\rho \in [-0.9, 0.9]$.
 - Encuentre el ρ que maximiza $\log(L(\rho))$ utilizando `argmax` (numpy).
7. Asumiendo la distribución del inciso 6, utilizando como coeficiente de correlación su estimación, hallar la recta de regresión. Graficar los datos con una nube de puntos y superponer la recta de regresión sobre ellos. Utilice la recta para estimar cuánto duraría la batería de un smartphone de 256GB de almacenamiento.
- 👁: Utilice `scatter` (matplotlib) para el gráfico.