

Proyecto Evaluativo Estadística. Tercer Año. Ciencias de la Computación.

Fase 1. Estadística Descriptiva, Estimación y Prueba de Hipótesis.

1. Genere una Población Normal** de tamaño 500, seleccione 8 muestras de tamaños varios (Mucho mayor que 30, mayor que 30, 30, 20), 4 muestras con remplazo y 4 sin remplazo.
 - a. Calcule para cada una de las muestras los Estadísticos Descriptivos, de la Conferencia 1.
 - b. Calcúlelos en la población inicial. Analice las diferencias.
 - c. Grafique los resultados****
 - d. Para cada muestra calcule los intervalos de confianza para la media y la varianza***
 - e. Analice las diferencias en los resultados de las muestras de tamaños similares.
2. De acuerdo a su set de datos*:
 - a. Utilice los Estadísticos Descriptivos estudiados en la Conferencia 1. Para describir el comportamiento de tres de sus variables. Seleccione las que sean más importantes y explique porque seleccionó estas.
 - b. Grafique los resultados***.
 - c. Interprete los resultados en términos del problema.
3. En la siguiente hoja aparecen 20 ejercicios. La orientación de la tercera pregunta es de acuerdo al número de cada equipo.

Hints.

* Es posible que necesite pre procesar los datos para buscar valores faltantes. Investigue las mejores formas de hacer esto.

**Genere primero una población y de ahí saque muestras usando la función `sample()`

*** Programe una función en R que calcule el intervalo y que reciba como parámetro una muestra.

**** Si necesita ayuda con las funciones de R o quiere modificar los gráficos consulte el libro *R into Action*, o el panel de ayuda del R Studio o el repositorio <https://cran.r-project.org>.

***** Si va a trabajar con variables categóricas se sugiere que las codifique.

Equipos:

1. ¿Hay diferencias significativas entre la cantidad de personas por debajo del límite de pobreza (NumUnderPov) en los estados de California (CA) y New Jersey (NJ)?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
2. ¿Se puede afirmar que en la concentración 0.799-0.201 el sensor QMC3 detecta en promedio valores mayores que el QM12?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
3. ¿Hay diferencias significativas entre los precios de venta de los teléfonos marca Apple y los Samsung?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
4. ¿Se puede afirmar que el promedio de carga de trabajo por día (Work load Average/day) es mayor en los meses de Junio que en los de Agosto?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
5. ¿Será el coeficiente prismático de la muestra menor que 0.5?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
6. ¿Se puede afirmar que la cantidad de juegos ganados por Djokovich en el Abierto de Estados Unidos difiere con respecto a los demás torneos?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal. Requiere un procesamiento extra de los datos.
7. ¿Se puede afirmar la varianza del flujo de combustible (Fuel Flow) de la muestra es mayor que 1 kg/s?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
8. ¿Será menor el promedio del área de los incendios ocurridos en el mes de agosto que los ocurridos en el mes de septiembre?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
9. ¿Existe diferencias en la duración promedio de las campañas que siguen las personas casadas contra las que siguen las personas solteras?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
10. ¿Hay diferencias significativas entre el promedio de años dedicados a la educación y la cantidad de ingreso de los censados?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
11. ¿Se puede afirmar que la varianza la agilidad de los jugadores es menor que 12 puntos?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.

12. ¿Será mayor el promedio de agilidad requerida en los deportes individuales que en los deportes por equipo?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal. Debe generar una variable nueva.
13. ¿Se puede afirmar que la media de la agilidad de los jugadores es mayor que 11 puntos?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
14. ¿Existe diferencias entre el precio del presupuesto de películas de acción y dramas?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
15. ¿Existe diferencia significativa entre el total de ganancias (total gross) en taquilla de las películas de la década de los 80 con las películas de la década de los 90?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
16. ¿Se puede afirmar que la media del máximo alcance vertical de los jugadores es mayor que 130?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
17. ¿Sera el promedio de aparición en el plato mayor qué 645?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
18. ¿Se puede afirmar que la media del alcance de pie, de los jugadores es mayor que 98?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
19. ¿Hay diferencias significativas entre el pH promedio de los vinos blancos con el ph promedio de los vinos tintos?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
20. ¿Se puede afirmar que la media del Sprint de los jugadores es igual a 3.5?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.
21. ¿Se puede afirmar que la media de tiempo de los 5000 m es superior a los 15 minutos?
Sugerencia: Asuma que todas las observaciones provienen de una distribución normal.