

Tarea 1

Estadística Descriptiva, Correlación y Regresión Lineal Simple de Primer Orden

Descripción del formato de entrega de la tarea (que tiene Parte 1 y Parte 2):

- Solamente un integrante de cada grupo deberá entregar la tarea en el curso EVA.
- Se deberá entregar un único documento en formato PDF (con el nombre de todos los integrantes del grupo) que será el informe de la tarea, con resultados (tablas, gráficos, etc.) e interpretaciones que respondan a cada ítem.
- Fecha máxima de entrega: Jueves 14/09 (23:59hs)

PARTE 1

Datos experimentales

En un experimento estadístico se trabaja con 50 registros proveniente de operaciones de cosecha, los datos se obtuvieron de una computadora abordo de un harvester:

- DAP(mm)
- volumen_cosechado_total(m3)
- tiempo_procesamiento(min)
- productividad(m3/hr)
- especie
- turno
- operador

Los datos se presentan en el archivo "datos_cosecha.csv".

Se pide:

- 1) Presentar alguna salida de Infostat que permita visualizar la media y los valores mínimos y máximos de las variables DAP, volumen_cosechado_total y tiempo_procesamiento. ¿Cuál es el tiempo promedio de las operaciones de cosecha?
- 2) Presentar una tabla de frecuencias para las variable productividad (dividir las productividades en rangos de 10 m³).
- 3) Presentar un diagrama de caja para visualizar el tiempo_procesamientos en la muestra. El gráfico debe ser de color verde y debe tener el título "Tiempo de procesamiento en minutos".
- 4) Realizar un gráfico de caja personalizado para visualizar el volumen cosechado, en función del operador. Interpretar los resultados.
- 5) Presentar un histograma de frecuencias relativas para el DAP de los registros. Se deberá presentar el gráfico con 6 clases y se deberá agregar el polígono de frecuencias uniendo las marcas de cada clase.
- 6) Realizar una tabla de frecuencias relativas para la variable tiempo_procesamiento, trabajando con 4 niveles. A partir de esa tabla presentar un diagrama de barras para la variable en función de los niveles.

PARTE 2

Datos experimentales

Se quiere analizar el efecto de la aplicación de fertilizante a 3 variedades de un cultivo de un cereal, sobre el rendimiento cosechado por hectárea. En el conjunto de datos se presentan los resultados de las mediciones de 50 muestras, de las variables: tratamiento, repetición, variedad del cereal, rendimiento por hectárea, y altura de la planta.

Los datos se presentan en el archivo "datos_exp.csv"

Se pide

1. Presentar un gráfico en el que se visualice el comportamiento conjunto de las variables Altura y Rendimiento, según la Variedad. Personalizar el color, tamaño y símbolo de cada serie de datos y cambiar el nombre de los ejes y del título. ¿Cuál es su opinión sobre el gráfico obtenido? (comentar brevemente).
2. Realizar un análisis de correlación lineal de Pearson para Altura y Rendimiento, trabajando primero con cada Variedad por separado y luego con todo el set de datos en conjunto. Presentar los resultados de estas pruebas y comentar los resultados. ¿Se corresponden los resultados con lo que se visualizaba en el gráfico realizado en el punto 1?
3. Se propone un análisis de regresión de primer orden para la variable Rendimiento, en función de las dosis de fertilizante aplicadas. Presentar la salida de INFOSAT que permite visualizar los resultados del análisis de la varianza, coeficientes de regresión e indicadores de bondad de ajuste.
4. Comprobar gráfica y analíticamente el cumplimiento de los supuestos del modelo propuesto en 3.