

Evaluación Sumativa N°3

Procesamiento y Análisis de Datos con R

Ponderación: 25%

Tipo: Grupal (En Parejas) opcional, es decir, puede desarrollarla de manera individual.

Fecha Entrega: 01-12-2025 - 23:00 hrs, Vía Plataforma

No se reciben trabajos por correo

1. Contexto del Caso

La Oficina de Estudios y Políticas Agrarias (ODEPA) es el organismo del Ministerio de Agricultura encargado de recopilar, analizar y difundir información estratégica sobre el sector silvoagropecuario de Chile.

Su misión es proporcionar información objetiva, oportuna y verificable que permita:

- Diseñar políticas agrícolas.
- Monitorear el comportamiento de los mercados.
- Mejorar la competitividad del sector.
- Apoyar decisiones de productores, comerciantes y autoridades.

En este marco, ODEPA mantiene series históricas con datos de frutas y hortalizas, incluyendo precios mayoristas, volúmenes comercializados, unidades de venta, zonas geográficas y origen de los productos.

Los estudiantes deberán procesar, analizar y visualizar estos datos utilizando R, generando un informe técnico individual.

Debe utilizar la información de 10 años, es decir, del 2016 al 2025

Link descarga: <https://datos.gob.cl/dataset/precios-mayoristas-de-frutas-y-hortalizas>

2. Objetivo de la Evaluación

Demostrar la capacidad para:

- Limpiar, transformar y preparar datos (3.1.1)
- Aplicar modelos estadísticos y de regresión (3.1.2)
- Organizar y justificar las decisiones metodológicas tomadas durante el análisis (3.1.5)
- Generar análisis comparativos y gráficos que aporten valor e interpretación

3. Tareas a Realizar

A. Proceso de Obtención y Limpieza de Datos (3.1.1)

- El estudiante deberá realizar un proceso completo de preparación.
- Dentro del informe debe explicar con detalle las limpiezas aplicadas.
- Se espera que consideren al menos estos elementos (sin limitarse a ellos):

1. Normalización de nombres de columnas

- Estándar, sin espacios, sin acentos, formato consistente.

2. Conversión de tipos

- Variables numéricas (precio, volumen).
- Variables categóricas (producto, unidad, origen, región).
- Conversión de fecha a formato correcto.

3. Generación de nuevas columnas (obligatorio)

- Año, mes, día
- Variable binaria “nacional/extranjero”
- Variación porcentual entre años
- Agrupador de categoría (fruta / hortaliza)
- Rangos o tramos de precios

4. Tratamiento de valores nulos

Justificar decisiones:

- Reemplazar
- Eliminar
- Imputar según criterios

5. Considerar la unidad de comercialización

Los datos no están en la misma unidad, por lo que deberán:

- Detectar distintas unidades (kg, docena, bandeja, caja, etc.)
- Elegir cómo tratarlas:

- Homogeneizar
- Analizar por unidad
- Separar series
- Reconvertir

Justificar su decisión en el informe.

6. Eliminación de caracteres o ruidos

Comas, símbolos, strings dentro de columnas numéricas, etc.

B. Análisis Estadístico y Exploratorio (sin escogerles variables)

El estudiante debe realizar un análisis profundo utilizando:

1. Estadística descriptiva básica
 - Distribuciones
 - Medias
 - Rangos
 - Percentiles
 - Tablas de frecuencia

C. Modelos de Regresión (obligatorio)

Se debe aplicar al menos un modelo de regresión, a elección del estudiante:

- Regresión lineal simple
- Regresión lineal múltiple
- Regresión polinómica
- Tendencia en series temporales

Debe explicar:

- Variables independientes usadas
- Variable dependiente
- Resultados
- Interpretación gráfica y escrita

D. Visualizaciones (mínimo 10 gráficos)

Los gráficos deben incluir interpretaciones, no solo ser mostrados.

Los tipos de visualización a considerar:

- Series de tiempo
- Comparación entre categorías
- Análisis por origen (nacional/extranjero)
- Comparación entre frutas y hortalizas

- Porcentajes
- Mapas o gráficos por región (si corresponde)
- Top N según criterio propio
- Regresión lineal con línea ajustada

Cada gráfico debe:

- ✓ Tener título
- ✓ Describir qué se interpreta
- ✓ Explicar el significado del resultado

E. Organización y Metodología (3.1.5)

En el informe se debe explicar:

- Qué decisiones metodológicas se tomaron y por qué
- Dificultades encontradas
- Cómo se ajustó el proceso para cumplir los objetivos
- Cómo se organizó el flujo de trabajo
- Qué mejoras se aplicarían en una versión futura

4. Estructura del Informe

Se evaluará que el informe contenga:

1. Portada
2. Introducción
3. Desarrollo
 - Limpieza
 - Transformaciones
 - Nuevas columnas
 - Análisis descriptivo
 - Inferencia estadística
 - Regresión
 - Visualizaciones
4. Conclusiones
5. Anexos (código completo)