

Sigla Asignatura	BIY7121	Nombre de la Asignatura	Minería de datos	Tiempo	2 horas
Experiencia de Aprendizaje N° 2	Genera información y conocimiento útil desde las diversas fuentes de datos para permitir la toma de decisiones de acuerdo con las necesidades de la organización.				
Actividad N° 2.5	2.5 Python y modelos estadísticos				
Nombre del Recurso Didáctico	2.5.4 Actividad statsmodels				

1. Aprendizajes e indicadores de logro

Aprendizajes (Procedimentales, Actitudinales y conceptuales)	Indicadores de logro
Aplicar metodologías para el desarrollo de proyectos minería de datos, considerando las herramientas de estadística básica para aplicar en este tipo de proyecto	Realiza la exploración de los datos utilizando herramientas de estadística básica para determinar los valores relevantes.
Analizar los datos mediante la aplicación de modelos estadísticos, utilizando el lenguaje Python.	Realiza la transformación de los datos, utilizando el framework Anaconda, clasificándolos de acuerdo a sus características.
	Aplica las métricas asociadas a estadística básica para detectar los valores relevantes.
	Aplica los modelos estadísticos en los datos utilizando el lenguaje definido.
	Grafica los resultados obtenidos a partir del trabajo con los datos, de acuerdo a las necesidades y características del cliente.
Es riguroso con el manejo de los datos para la discriminación de los datos procesados.	Es capaz de contextualizar los resultados obtenidos de los modelos estadísticos en el entorno organizacional.
	Sigue los procedimientos en la aplicación de los modelos estadísticos.

2. Descripción general actividad

Esta actividad tiene carácter formativo con el objetivo de aplicar el uso de statsmodels.

Se debe desarrollar en grupos de 3 o 4 estudiantes.

Disponen de 1 sesión para resolver la actividad. Al finalizar deberá presentar un informe con el resultado de los ejercicios propuestos y su desarrollo completo.

OBJETIVO

Conocer los aspectos generales de la relación entre las variables que forman parte de un conjunto de datos.

Ejercicio

Considerando el conjunto de datos disponible en el archivo “2.5.5 Salary_data.csv”. Este conjunto de datos cuenta con la información de 30 colaboradores, se tienen los años de experiencia y el salario de cada uno de ellos.

El objetivo es saber si los años de experiencia de los 30 colaboradores se encuentran asociados con el salario.

Se pide construir un notebook que cumpla con los siguientes requerimientos:

1. Carga de datos en un dataframe
2. Obtener las medidas estadísticas de las 2 características que se incluyen en el conjunto de datos:
 - a. Frecuencia (total de observaciones)
 - b. Promedio
 - c. Desviación estándar
 - d. Valor mínimo
 - e. Valor máximo
 - f. Los 3 primeros cuartiles
 - g. Valor máximo
3. Usando el método de los mínimos cuadrados (OLS) obtener un resumen del modelo.
4. Interpretar los resultados obtenidos en el paso anterior
5. Generar gráficos para:
 - a. Ecuación de regresión
 - b. Gráfico de pronósticos
 - c. Gráfico de errores (residuos)