

Trabajo de estadística matemáticas

El trabajo debe realizarlo utilizando Python. Enviar antes de las 2pm del 06/07/2020 por correo el archivo .py y un archivo Word con el análisis de cada una de las preguntas.

Regresión lineal

El archivo datos1.csv incluye 711 ejemplos de beneficiarios inscritos en un plan de seguro indicando las características del paciente, así como el total de gastos médicos cargados (charges) al plan para el año calendario. Las características son:

Edad (age): un número entero que indica la edad del beneficiario principal (excluyendo los mayores de 64 años, ya que generalmente están cubiertos por el gobierno de Estados Unidos).

Sexo (sex): el sexo del titular de la póliza, ya sea masculino o femenino.

Bmi: el índice de masa corporal. El bmi es igual al peso (en kilogramos) divididos por altura (en metros) al cuadrado. Un bmi ideal está en un rango de 18.5 a 24.9.

Hijos (children): un número entero que indica el número de hijos dependientes cubiertos por el plan de seguro.

Fumador (smoker): una variable categórica que indica si el asegurado fuma regularmente.

Región: el lugar de residencia del beneficiario en los Estados Unidos, dividido en cuatro regiones geográficas: noreste, sureste, suroeste o noroeste.

Observe que estas variables pueden estar relacionadas con las facturas de gastos médicos. Por ejemplo, podríamos esperar que las personas mayores y los fumadores estén en mayor riesgo de grandes gastos médicos.

1. Realice una exploración sobre los datos (estadísticas descriptivas).
2. Calcule la relación entre las variables (matriz de correlaciones).
3. Visualice la relación entre las variables (gráficos).
4. Contraste la hipótesis de que en las 4 regiones se tiene en promedio el mismo índice de masa corporal bmi (Análisis de varianzas).
5. Obtenga un modelo lineal que pueda predecir el total de gastos médicos cargados para un beneficiario.
6. Contraste la hipótesis. Donde son los coeficientes del modelo.
7. Calcule el coeficiente de determinación.
8. Realice un análisis de los residuos.

Análisis de Varianza.

Realice un programa que para resolver problemas de análisis de varianza. Elija un ejemplo ilustrativo y realice el análisis.