**Imagen que contiene Logotipo

Descripción generada automáticamenteLogotipo

Descripción generada automáticamenteINSTITUTO POLITÉCNICO  
NACIONAL**

**ESCUELA SUPERIOR  
DE CÓMPUTO**

Materia:

Data Mining

Profesora:

Fabiola Ocampo Botello

Tarea #6:

Regresión Lineal

Equipos #10 y #4:

* Alanís Garduño Mireya
* Ríos Rebollar Víctor Hugo
* García Cruz Octavio Arturo
* Sampayo Hernández Mauro
* Flores Ponce Alan Marcelo

Grupo:

3CV15

**EJERCICIO DE REGRESIÓN LINEAL. GRUPO 3CV15**

**Materia: Minería de datos Periodo escolar: 2023-2**

|  |
| --- |
| Equipo: **4** |

Nombre de los integrantes del equipo:

|  |  |
| --- | --- |
| 1) | Flores Ponce Alan Marcelo |
| 2) | García Cruz Octavio Arturo |
| 3) | Sampayo Hernández Mauro |

Una compañía aplica a sus vendedores en periodo de capacitación una prueba de ventas antes de salir a trabajar. La administración de la compañía está interesada en determinar la relación entre las calificaciones de la prueba y las ventas logradas por esos vendedores al final de un año de trabajo. Se recolectaron los siguientes datos de 10 agentes de ventas que han estado en el campo un año.

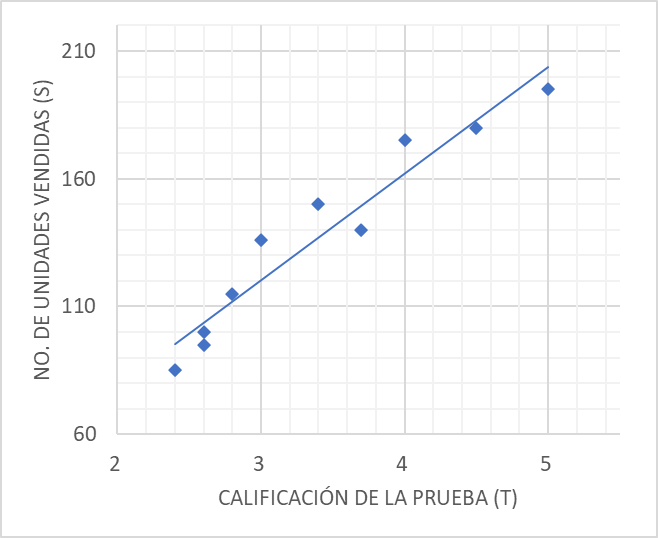
Se desea crear un modelo matemático que represente la relación de los datos. Utilice la guía proporcionada.

Ejercicio adaptado de Levín, Rubín, Balderas, Del Valle y Gómez. (2004). Estadística para administración y economía. Séptima Edición. Prentice-Hall.

|  |  |  |
| --- | --- | --- |
| **No. Vendedor** | **Calificación de la prueba (T)** | **No. De unidades vendidas (s)** |
| 1 | 2.6 | 95 |
| 2 | 3.7 | 140 |
| 3 | 2.4 | 85 |
| 4 | 4.5 | 180 |
| 5 | 2.6 | 100 |
| 6 | 5.0 | 195 |
| 7 | 2.8 | 115 |
| 8 | 3.0 | 136 |
| 9 | 4.0 | 175 |
| 10 | 3.4 | 150 |

Responder cada uno de los siguientes incisos. Agregar la generación de tablas de cálculos y la presentación de las fórmulas que utilice en cada sección.

Las tablas de los cálculos las debe de agregar en este reporte y enviar el archivo respectivo.

1. Generar la gráfica de variables

|  |  |
| --- | --- |
| Calificación de la prueba (T) | No. de unidades vendidas (S) |
| 2.6 | 95 |
| 3.7 | 140 |
| 2.4 | 85 |
| 4.5 | 180 |
| 2.6 | 100 |
| 5 | 195 |
| 2.8 | 115 |
| 3 | 136 |
| 4 | 175 |
| 3.4 | 150 |

1. Realice los cálculos *pasos a paso* para generar la ecuación de regresión.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Calificación de la prueba () | No. de unidades vendidas () |  |  |  |  |
| 2.6 | 95 | -0.8 | -42.1 | 33.68 | 0.64 |
| 3.7 | 140 | 0.3 | 2.9 | 0.87 | 0.09 |
| 2.4 | 85 | -1 | -52.1 | 52.1 | 1 |
| 4.5 | 180 | 1.1 | 42.9 | 47.19 | 1.21 |
| 2.6 | 100 | -0.8 | -37.1 | 29.68 | 0.64 |
| 5 | 195 | 1.6 | 57.9 | 92.64 | 2.56 |
| 2.8 | 115 | -0.6 | -22.1 | 13.26 | 0.36 |
| 3 | 136 | -0.4 | -1.1 | 0.44 | 0.16 |
| 4 | 175 | 0.6 | 37.9 | 22.74 | 0.36 |
| 3.4 | 150 | 0 | 12.9 | 0 | 0 |
|  |  |  |  |  |  |

**Ecuación de Regresión:**

1. Realice la verificación de la ecuación de regresión de una recta generada con el método de mínimos cuadrados.

|  |  |  |  |
| --- | --- | --- | --- |
| **Calificación de la prueba ()** | **No. de unidades vendidas ()** |  | **Error** |
| 2.6 | 95 | 103.7556 | -8.7556 |
| 3.7 | 140 | 149.6047 | -9.6047 |
| 2.4 | 85 | 95.4194 | -10.4194 |
| 4.5 | 180 | 182.9495 | -2.9495 |
| 2.6 | 100 | 103.7556 | -3.7556 |
| 5 | 195 | 203.79 | -8.79 |
| 2.8 | 115 | 112.0918 | 2.9082 |
| 3 | 136 | 120.428 | 15.572 |
| 4 | 175 | 162.109 | 12.891 |
| 3.4 | 150 | 137.1004 | 12.8996 |
|  |  |  | **-0.004** |

1. Realice los siguientes cálculos (muestre el proceso)
2. Suma de cuadrados debida al error

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Calificación de la prueba ()** | **No. de unidades vendidas ()** |  | **Error** | **Error al cuadrado** |
| 2.6 | 95 | 103.7556 | -8.7556 | 76.66053136 |
| 3.7 | 140 | 149.6047 | -9.6047 | 92.25026209 |
| 2.4 | 85 | 95.4194 | -10.4194 | 108.5638964 |
| 4.5 | 180 | 182.9495 | -2.9495 | 8.69955025 |
| 2.6 | 100 | 103.7556 | -3.7556 | 14.10453136 |
| 5 | 195 | 203.79 | -8.79 | 77.2641 |
| 2.8 | 115 | 112.0918 | 2.9082 | 8.45762724 |
| 3 | 136 | 120.428 | 15.572 | 242.487184 |
| 4 | 175 | 162.109 | 12.891 | 166.177881 |
| 3.4 | 150 | 137.1004 | 12.8996 | 166.3996802 |

1. Suma total de cuadrados

|  |  |  |  |
| --- | --- | --- | --- |
| **Calificación de la prueba ()** | **No. de unidades vendidas ()** | **Desviación** | **Desviación al cuadrado** |
| 2.6 | 95 | -42.1 | 1772.41 |
| 3.7 | 140 | 2.9 | 8.41 |
| 2.4 | 85 | -52.1 | 2714.41 |
| 4.5 | 180 | 42.9 | 1840.41 |
| 2.6 | 100 | -37.1 | 1376.41 |
| 5 | 195 | 57.9 | 3352.41 |
| 2.8 | 115 | -22.1 | 488.41 |
| 3 | 136 | -1.1 | 1.21 |
| 4 | 175 | 37.9 | 1436.41 |
| 3.4 | 150 | 12.9 | 166.41 |

1. Suma de cuadrados debida a la regresión

|  |  |  |  |
| --- | --- | --- | --- |
| **Calificación de la prueba ()** | **No. de unidades vendidas ()** |  |  |
| 2.6 | 95 | 103.76 | 1111.85 |
| 3.7 | 140 | 149.60 | 156.37 |
| 2.4 | 85 | 95.42 | 1737.27 |
| 4.5 | 180 | 182.95 | 2102.18 |
| 2.6 | 100 | 103.76 | 1111.85 |
| 5 | 195 | 203.79 | 4447.56 |
| 2.8 | 115 | 112.09 | 625.41 |
| 3 | 136 | 120.43 | 277.96 |
| 4 | 175 | 162.11 | 625.45 |
| 3.4 | 150 | 137.10 | 0.00 |

1. El coeficiente de determinación
2. Exprese el significado del coeficiente de determinación encontrado

*Tenemos el 92.69% de bondad de ajuste de la ecuación de regresión estimada.*

1. El coeficiente de correlación y su significado

*Existe una correlación positiva, los valores de una variable tienden a incrementarse mientras que los valores de la otra variable descienden, y que por lo tanto existe una relación lineal fuerte entre x y y.*

1. Calcule los errores estándar de la estimación
2. Los intervalos de confianza

**Método 1:**

Sabiendo que la ecuación de la predicción es:

Supongamos 4 interrupciones.

Se estima que la clasificación en la prueba va a tener un valor de 162.10

El error estándar calculado es de

Para una confianza del 95% es necesario obtener 2 grados de libertad, por lo tanto:

Por ende, los intervalos de confianza son de 184.02 a 140.18 respecto a la clasificación de la prueba de honestidad

**Método 2:**

Tomando los valores además de 2 grados de libertad y haciendo uso de Ŷ 𝑆 la siguiente tabla, 𝑒

se harán los cálculos correspondientes.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Grados de Libertad | 0.75 | 0.9 | 0.95 | 0.97 | 0.99 | 0.995 | 0.9995 |
| 1 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 636.619 |
| 2 | 0.816 | 1.886 | 2.9620 | 4.303 | 6.965 | 9.925 | 31.598 |
| 3 | 0.765 | 1.6368 | 2.353 | 3.182 | 4.541 | 5.841 | 12.941 |
| 4 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 8.610 |

Mediante las siguientes fórmulas se van a calcular los intervalos de confianza con una confianza del 0.95 ó 95%

Al resolver las ecuaciones, se obtiene que la clasificación se encontrará 194.563 y 129.636con una confianza del 95%.

1. Aplique la prueba t para determinar si el modelo es estadísticamente significativo

Se generan las siguientes hipótesis considerando el parámetro

(No existe relación lineal entre la “calificación de la prueba” y el “no. de unidades vendidas”)

(Si la “calificación de la prueba” y el “no. de unidades vendidas” están relacionadas linealmente)

Se plantea el siguiente modelo matemático:

Sea:

Se tiene:

10.07545

El valor de *t > 0*, por lo que se puede concluir que **la “calificación de la prueba” y el “no. de unidades vendidas” están relacionadas linealmente**, lo que significa que el modelo es **estadísticamente significativo.**

1. Genere la ecuación de recta en el Knime incorporando prueba de normalidad y gráfico de residuales.

A picture containing text, screenshot, rectangle, diagram

Description automatically generated

**Imagen 1 Gráfica de Variables**

A screenshot of a computer

Description automatically generated with low confidence

**Imagen 2 Resultados de Regresión**

A picture containing line, text, parallel, diagram

Description automatically generated

**Imagen 3 Ecuación de la Recta**

**Texto, Tabla

Descripción generada automáticamente**

**Imagen 4 Prueba de Normalidad**

**A screenshot of a computer

Description automatically generated with low confidence**

**Imagen 5 Residuos**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**A picture containing text, screenshot, rectangle, diagram

Description automatically generated**

**Imagen 6 y 7 Gráficos de Variables y Residuales**