# Imagen que contiene edificio, bebidas, exterior, alimentos Descripción generada automáticamente4. Actividad en clase de Regresión múltiple

1. **SERVICIO VENDING MACHINES** John comenzó a trabajar para DHL como especialista logístico. Después de hacer un trabajo impresionante, recibió una oferta por un puesto en FEMSA Logistics. FEMSA tuvo una relación especial con el Tec de Monterrey, su alma mater. Estaba feliz de aceptarlo.

En su primera asignación en su nuevo trabajo, necesita analizar las rutas de servicio de máquinas expendedoras en el sistema de distribución de botellas y latas de Coca-Cola. Está interesado en predecir el tiempo requerido por el conductor de la ruta para dar servicio a las máquinas expendedoras.

Fue un problema interesante, ya que algunas máquinas expendedoras estaban cerca de la calle dentro de pequeñas tiendas de esquina, pero otras estaban lejos de la calle dentro de edificios de oficinas, campus universitarios o centros comerciales.

Imagen que contiene persona, hombre, tabla, viendo

Descripción generada automáticamenteDespués de hacer algunas preguntas a los conductores de los camiones repartidores, piensa que las dos variables más importantes que afectan los tiempos de entrega son el número de casos de producto abastecido y la distancia recorrida por el conductor de la ruta.

Tenía sentido. Cuantas más cajas se almacenaban en el carro de mano, más difícil era moverlo. Cuanto mayor sea la distancia para caminar, más tiempo se requiere para llegar a la máquina expendedora.

John pidió a dos conductores de ruta que registraran cuánto tiempo necesitaban para dar servicio a las máquinas expendedoras registrando cuántos casos almacenaban y hasta dónde caminaban.

**¿Hay alguna relación entre las cajas de producto almacenado y la distancia recorrida por el conductor de la ruta?**

**¿Es posible predecir el tiempo necesario para dar servicio a las máquinas expendedoras?**

**Argumente ampliamente**

1. Estimar e interpretar los parámetros que forman parte del Modelo de Regresión Lineal Múltiple.
2. Calcular e interpretar algunos indicadores de la calidad del modelo, como el coeficiente de determinación ajustado (R^2) y el error estándar.

|  |  |
| --- | --- |
| *Estadísticas de la regresión* | |
| Coeficiente de correlación múltiple | 0.97958856 |
| Coeficiente de determinación R^2 | 0.95959375 |
| R^2 ajustado | 0.95592045 |
| Error típico | 3.25947345 |
| Observaciones | 25 |

La calidad del modelo según la tabla, el r^2 ajustado es de 0.955 y mientras más cercano esté al 1, el modelo es más exacto. Por otro lado, el error típico es 3.25 que comparando los valores del tiempo (y) es una magnitud pequeña. Lo que indica que las observaciones están muy cercanas a la línea ajustada

1. Realizar el ANOVA para determinar la significancia del modelo de regresión.

* Ho El modelo no tiene un buen ajuste
* Ha El modelo tiene un buen ajuste

Regla de decisión: Rechazar Ho si el valor crítico de F es menor que 0.05

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANÁLISIS DE VARIANZA | |  |  |  |  |
|  | *Grados de libertad* | *Suma de cuadrados* | *Promedio de los cuadrados* | *F* | *Valor crítico de F* |
| Regresión | 2 | 5550.81092 | 2775.40546 | 261.235109 | 0.0000000000000004687 |
| Residuos | 22 | 233.731677 | 10.6241672 |  |  |
| Total | 24 | 5784.5426 |  |  |  |

Rechazo Ho y declaro que tiene un bien ajuste debido a que su valor crítico de F es de : 0.0000000000000004687

1. Validar los parámetros del modelo aplicando la prueba t.

* Ho: b1 = 0
* Ho: b1 =! 0
* Ho: b2 = 0
* Ho: b2 =! 0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Coeficientes* | *Error típico* | *Estadístico t* | *Probabilidad* |
| Intercepción | 2.34123115 | 1.09673017 | 2.13473762 | 0.04417012 |
| Cajas | 1.61590721 | 0.17073492 | 9.46442138 | 3.2549E-09 |
| Distancia (metros) | 0.01438483 | 0.00361309 | 3.98131314 | 0.00063125 |

Rechazo Ho porque el valor de probabilidad es muy pequeño y las variables x1 y x2 (cajas y distancia) son unos buenos predictores del tiempo de servicio a las vending machines.

1. Realice una estimación del tiempo que tardará en darle servicio a un vending machine un trabajador que lleve 9 cajas y tenga que desplazarse 700 metros.

Tiempo = 2.34123115+ 1.61590721\* cajas +0.01438483 \*metros

Tiempo = 2.34123115+ 1.61590721\*(9 cajas) +0.01438483 (700 metros) =

26.9537744 minutos

1. Brinde una respuesta a las preguntas de investigación planteadas.

**¿Hay alguna relación entre las cajas de producto almacenado y la distancia recorrida por el conductor de la ruta?**

**¿Es posible predecir el tiempo necesario para dar servicio a las máquinas expendedoras?**

Sí porque tanto los errores cuadrados mínimos de los datos como el error típico resultaron con buenas métricas (R^2 muy cercano a 1 y los errores típicos del orden de los 3 min). También, ha pasado la prueba de hipótesis de validez del modelo y de los coeficientes con un valor p<0.05.

Se pudo predecir que se tarda casi media hora en dar servicio a una máquina que requiere 9 cajas y está a 700 metros de distancia.

|  |  |  |  |
| --- | --- | --- | --- |
| **Obs** | **Tiempo (min)** | **Cajas** | **Distancia (metros)** |
| 1 | 16.68 | 7 | 560 |
| 2 | 11.5 | 3 | 220 |
| 3 | 12.03 | 3 | 340 |
| 4 | 14.88 | 4 | 80 |
| 5 | 13.75 | 6 | 150 |
| 6 | 18.11 | 7 | 330 |
| 7 | 8 | 2 | 110 |
| 8 | 17.83 | 7 | 210 |
| 9 | 79.24 | 30 | 1460 |
| 10 | 21.5 | 5 | 605 |
| 11 | 40.33 | 16 | 688 |
| 12 | 21 | 10 | 215 |
| 13 | 13.5 | 4 | 255 |
| 14 | 19.75 | 6 | 462 |
| 15 | 24 | 9 | 448 |
| 16 | 29 | 10 | 776 |
| 17 | 15.35 | 6 | 200 |
| 18 | 19 | 7 | 132 |
| 19 | 9.5 | 3 | 36 |
| 20 | 35.1 | 17 | 770 |
| 21 | 17.9 | 10 | 140 |
| 22 | 52.32 | 26 | 810 |
| 23 | 18.75 | 9 | 450 |
| 24 | 19.83 | 8 | 635 |
| 25 | 10.75 | 4 | 150 |

1. **ESTIMACIÓN DE LA DUREZA DEL ACERO**. Jorge trabaja para AHMSA, Altos Hornos de México, una de las primeras empresas siderúrgicas en México, que data de 1941.

AMHSA está planeando producir acero en frío de lámina reducida a partir de losas que consisten en 0.15 por ciento de cobre a una temperatura de recocido de 1,150 grados Fahrenheit.

Antes de que comience la producción, Jorge necesita estimar la dureza promedio (Rockwell 30-T) que tendrá la losa de acero. Esto le ayudará a establecer algunos parámetros del proceso de rodadura.

Para determinar esto, Jorge recogió datos sobre 10 muestras diferentes de losas de acero con diferentes contenidos de cobre y temperaturas de recocido.

**Ayude a Jorge a estimar la dureza esperada de las láminas que contienen 0.15 por ciento de cobre a una temperatura de recocido de 1,150 F**.

1. Estimar e interpretar los parámetros que forman parte del Modelo de Regresión Lineal Múltiple.
2. Calcular e interpretar algunos indicadores de la calidad del modelo, como el coeficiente de determinación ajustado (R^2) y el error estándar.
3. Realizar el ANOVA para determinar la significancia del modelo de regresión.
4. Validar los parámetros del modelo aplicando la prueba t.
5. **Ayude a Jorge a estimar la dureza esperada de las láminas que contienen 0.15 por ciento de cobre a una temperatura de recocido de 1,150 F**

|  |  |  |  |
| --- | --- | --- | --- |
| Obs | Dureza | Contenido de cobre | temperatura de recocido |
| 1 | 79.2 | 0.02 | 1050 |
| 2 | 64 | 0.03 | 1200 |
| 3 | 55.7 | 0.03 | 1250 |
| 4 | 56.3 | 0.04 | 1300 |
| 5 | 58.6 | 0.1 | 1300 |
| 6 | 84.3 | 0.15 | 1000 |
| 7 | 70.4 | 0.15 | 1100 |
| 8 | 61.3 | 0.09 | 1200 |
| 9 | 51.3 | 0.13 | 1400 |
| 10 | 49.8 | 0.09 | 1400 |