TERCER PARCIAL DE INFERENCIA ESTADISTICA

GRUPO-11

PROFESOR: NORBEY MARIN ARREDONDO

**INFORMACIÓN IMPORTANTE, POR FAVOR LEER ENTES DE COMENZAR EL EXAMEN:**

**- Para todas las pruebas utilice un nivel de significancia del 5%.**

**- Durante la presentación del examen, usted debe tener apagado y guardado en su maletín el celular. Si usted es sorprendido infringiendo esta restricción, su examen le será anulado.**

**- No se permite el intercambio de portátiles entre compañeros**

**- Si usted es sorprendido dando copia o intentando copiar, el examen será anulado**

**- Abra espacios debajo de cada item para que coloque allí: la salida de R y la interpretación del resultado. Solo se acepta como respuesta la información que está debajo de cada item.**

**- Construya un archivo Word de informe con las salidas de R y la interpretación de los resultados. Si la salida de R no aparece, su respuesta no tiene validez. Si quiere, guárdelo como pdf.**

**- La interpretación del resultado tiene un valor de 50% del punto.**

**- El archivo debe ser enviado al sistema INTU a más tardar a las 4 pm, a esa hora se cierra el sistema. No se califican archivos enviados al correo.**

**- Se califica el archivo que sea enviado al sistema INTU, revise y este seguro del archivo a enviar. Solo se acepta un archivo por estudiante.**

**- No se prestará asesoría durante la duración del examen**

1. Una compañía de transportes se encuentra evaluando la forma de incrementar el rendimiento del combustible a sus vehículos, con el propósito de disminuir los gastos de operación. Desarrolla un experimento tomando un solo carro, elige un solo conductor para que maneje siempre a la misma velocidad, en un mismo recorrido. El carro se tanquea siempre con la misma cantidad de gasolina, y se conduce hasta que esta se termine. Primero se desarrollan las pruebas sin agregar aditivo para poder tener un punto de referencia. Luego se adiciona a la gasolina un diferente tipo de aditivo (A, B, C, y D) y se realiza nuevamente el recorrido. Cuando se agota la gasolina se registra el número de millas totales recorridas desde el momento del tanqueo. Los resultados del experimento se presentan en la siguiente tabla:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sin aditivo | Con aditivo A | Con aditivo B | Con aditivo C | Con aditivo D |
| 18.6 | 25.5 | 32.8 | 30.5 | 21.8 |
| 24.5 | 29.5 | 26.4 | 29.8 | 19.5 |
| 15.4 | 30.2 | 22.4 | 25.6 | 20.3 |
| 25.2 | 26.4 | 28.6 | 28.2 | 18.6 |
| 28.6 | 24.3 | 19.8 | 32.8 | 21.5 |
| 18.3 | 30.5 | 20.5 | 22.6 | 25.6 |

a. Con la información del problema identifique: **(0.5)**

- el objetivo del experimento

El objetivo principal de este estudio es analizar cómo la adición de diferentes aditivos en la gasolina influye en el rendimiento del combustible de un vehículo. Se pretende identificar cuál de los aditivos permite al vehículo recorrer una mayor distancia con la misma cantidad de gasolina, lo que, en consecuencia, reduciría los gastos de operación de la empresa de transportes.

- el factor de estudio y sus niveles

El factor de estudio en este caso es el tipo de aditivo que se mezcla con la gasolina. Los niveles para este factor son los siguientes:

1. Sin aditivo
2. Aditivo A
3. Aditivo B
4. Aditivo C
5. Aditivo D

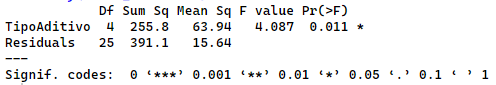
- la variable de respuesta. Explique adecuadamente si esta se maximiza o minimiza.

La variable de respuesta es el número de millas recorridas por el vehículo usando una cantidad fija de gasolina. Esta variable se maximiza, ya que el propósito es que el vehículo logre recorrer la mayor cantidad de millas posible, lo que aumentaría la eficiencia del combustible y ayudaría a reducir los costos operativos de la compañía.

b. Se puede asegurar que hay diferencia en el rendimiento promedio de la gasolina entre los grupos evaluados? Plantee, desarrolle e interprete adecuadamente la hipótesis. **(0.7)**

**H₁: Al menos un rendimiento promedio es diferente.**

**H₀: μ\_sin aditivo = μ\_A = μ\_B = μ\_C = μ\_D**

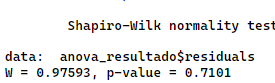
****

**Los resultados del análisis ANOVA muestran que el factor "Aditivo" presenta un valor de F de 4.087 y un valor p de 0.011, indicando una diferencia estadísticamente significativa en el rendimiento promedio de la gasolina entre los grupos (p < 0.05). Esto significa que, al comparar los distintos aditivos de combustible (y el grupo sin aditivo), se observa un rendimiento diferente en las millas recorridas. Por lo tanto, es posible concluir que al menos uno de los aditivos (o la falta de aditivo) afecta significativamente el rendimiento de la gasolina.**

c. Desarrolle las pruebas de supuestos del modelo. Plantee, desarrolle e interprete adecuadamente las hipótesis. **(0.6)**

**Evaluación de Normalidad**

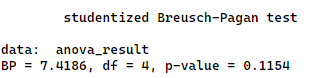
* **H₀: Los residuos del modelo se distribuyen de manera normal.**
* **H₁: La distribución de los residuos del modelo ANOVA no es normal.**



El valor p de la prueba de normalidad de Shapiro-Wilk es 0.7101, lo que es mayor que 0.05. Esto significa que no se rechaza la hipótesis nula de que los residuos siguen una distribución normal. Por lo tanto, podemos afirmar que el modelo cumple con el supuesto de normalidad de los residuos.

**Prueba de Breusch-Pagan para Homogeneidad de Varianzas**

* **H₀: Las varianzas de los errores son constantes a través de los diferentes niveles del grupo.**
* **H₁: Las varianzas de los errores no son constantes y varían entre los diferentes niveles del grupo.**

****

**El valor p de la prueba de Breusch-Pagan es 0.1154, lo cual es superior a 0.05. Esto implica que no podemos rechazar la hipótesis nula, lo que sugiere que las varianzas de los residuos son homogéneas entre los grupos evaluados. En conclusión, el modelo cumple con el supuesto de homocedasticidad y no se observa evidencia de heterocedasticidad.**

**Verificación de Autocorrelación en los Residuos**

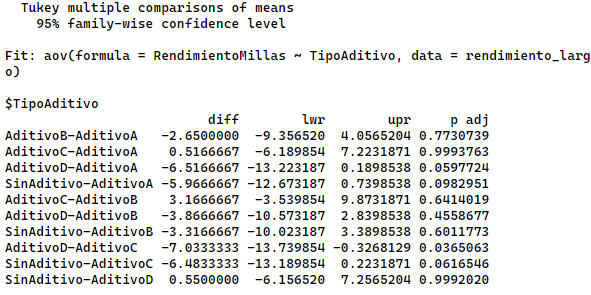
* **H₀: No existe autocorrelación significativa entre los residuos del modelo.**
* **H₁: Los residuos presentan autocorrelación significativa.**

****

**El estadístico de Durbin-Watson es 2.2614, con un p-valor de 0.41. Esto implica que no hay autocorrelación significativa en los residuos, ya que el p-valor es mayor que 0.05. Por lo tanto, los residuos son independientes, lo que cumple con el supuesto de independencia en el modelo.**

d. Determine el (los) tipo(s) de aditivo que favorece el rendimiento del combustible. Plantee, desarrolle e interprete adecuadamente las hipótesis. **(0.7)**

* **H₀ No existen diferencias significativas en el rendimiento de las millas entre los distintos tipos de aditivos.**
* **H₁ Existen diferencias significativas en el rendimiento de las millas entre los distintos tipos de aditivos.**



2. La empresa de Licores del Valle considera que las bajas ventas de los productos en la ciudad de Cali están siendo afectadas por la importación de todo tipo de licores tanto nacionales como extranjeros. El gerente solicita a sus asesores que mediante un modelo de regresión lineal simple relacionen como las importaciones están afectando las ventas. El encargado toma los registros de las ventas y las importaciones semanales en millones de pesos para los últimos cuatro meses. Los datos se registran a continuación: **(2.5)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ventas  Licor local | 559.4 | 425.2 | 388.4 | 285.8 | 253.8 | 195.3 | 135.5 | 132.8 | 120.8 | 125.7 | 120.8 | 115.8 | 380.6 | 172.5 | 160.5 | 116.2 |
| Importaciones  de licor | 156 | 187 | 199 | 212 | 215 | 250 | 280 | 282 | 295 | 312 | 326 | 355 | 125 | 270 | 265 | 322 |

a. Determine cuál es la variable dependiente y cual la independiente. Explique porque las considera así.

Texto

Descripción generada automáticamente

* **Variable dependiente**: Volumen de Ventas de Licores Locales
* **Variable independiente**: Importaciones de Licores

**Explicación**: En este caso, el **volumen de ventas de licores locales** se considera la variable dependiente, dado que es el fenómeno que se desea investigar en función de las importaciones. Las **importaciones de licores** son la variable independiente, pues se cree que tienen una relación inversa con las ventas de los productos locales, y buscamos medir este impacto.

b. Determine el coeficiente de correlación lineal. Plantee, desarrolle e interprete adecuadamente la hipótesis sobre la correlación.

**H0​**: No existe una correlación significativa entre importaciones y ventas de licor

**H1​**: Existe una correlación significativa entre importaciones y ventas.



Con un coeficiente de correlación de **-0.8951**, observamos una **fuerte relación negativa** entre el volumen de importaciones y las ventas de licor local. Esto significa que a mayor importación de licor, se presenta una notable disminución en las ventas locales, lo cual puede reflejar una competencia desfavorable para el licor local debido al incremento de productos importados. El valor alto en negativo refuerza la hipótesis de que existe un impacto negativo considerable de las importaciones sobre las ventas en el mercado.

c. Construya el modelo de regresión ajustado. Interprete adecuadamente los parámetros del modelo.

Texto

Descripción generada automáticamente

**Modelo de regresión:**

**Ventas\_Cali=710.29−1.89×Importaciones**

Interpretación:

Intercepto (710.2885): Este valor implica que, sin importaciones, el mercado de licor local podría generar ventas de alrededor de 710.29 millones de pesos. Esto refuerza la idea de una base fuerte de ventas locales en un contexto sin productos importados.

Pendiente (-1.8947): La pendiente indica que por cada aumento de un millón de pesos en las importaciones, las ventas de licor local disminuyen en aproximadamente 1.89 millones de pesos. Esto sugiere que el crecimiento de las importaciones está relacionado con una caída proporcional en las ventas de licores locales, afectando negativamente a la empresa en la región.

d. Plantee y desarrolle la prueba de hipótesis para comprobar si el modelo pasa por el origen. Explique claramente la conclusión.

**Hipótesis para el análisis**:

* H0: El intercepto es cero, indicando que el modelo debe pasar por el origen.
* H1​: El intercepto es diferente de cero, por lo que el modelo no pasa por el origen.

**Evaluación: La estimación del intercepto es de 710.2885, con un valor p extremadamente bajo (3.61e-08), lo cual es suficiente para rechazar H0​ al nivel de significancia de 0.05.**

**Conclusión final: Dado el rechazo de H0, confirmamos que el modelo no pasa por el origen. Esto implica que, en ausencia de importaciones, se espera que las ventas locales tengan un valor positivo significativo.**

e. Plantee y desarrolle la prueba de hipótesis de linealidad del modelo.

**Hipótesis**:

* **H0​**: La pendiente es igual a cero, lo cual sugiere que no hay relación lineal entre importaciones y ventas locales.
* **H1​**: La pendiente es distinta de cero, indicando una relación lineal significativa entre las variables.

**Evaluación: Con un valor ppp de 2.83e-06 para la pendiente, que es mucho menor a 0.05, rechazamos H0. Esto significa que la pendiente es significativa.**

**Conclusión: La evidencia respalda una relación lineal significativa entre las importaciones y las ventas locales, validando el uso de un modelo lineal para describir esta relación.**

f. Determine e interprete adecuadamente el coeficiente de determinación del modelo.

**Se obtuvo en el resumen del modelo (punto c)**

**R2**: El coeficiente de determinación obtenido es **0.8012**.

**Interpretación**: Este valor de R2 significa que el modelo explica alrededor del 80.12% de la variación en las ventas locales de licor debido a las importaciones. Esto sugiere que el modelo de regresión tiene un ajuste sólido, ya que una gran parte de la variabilidad en las ventas puede atribuirse a los niveles de importación. Un R2 tan alto indica que el modelo lineal es adecuado para estudiar esta relación.

g. Plantee, desarrolle e interprete adecuadamente las hipótesis de cada uno de los supuestos del modelo de regresión.

**1. Supuesto de Independencia de los Residuos**

* **Hipótesis**:
  + **H0​**: Los residuos son independientes y no presentan autocorrelación.
  + **H1​**: Los residuos están autocorrelacionados.

**2. Supuesto de Homoscedasticidad**

* **Hipótesis**:
  + **H1​**: La varianza de los residuos es homogénea.
  + **H1​**: Los residuos presentan varianza heterogénea.

**3. Supuesto de Normalidad de los Residuos**

* **Hipótesis**:
  + **H0​**: Los residuos se distribuyen de forma normal.
  + **H1**: Los residuos no siguen una distribución normal.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**1. Supuesto de Independencia de los Residuos**

* **Prueba de Durbin-Watson**: Con DW = **0.97334** y un p-valor de **0.007263**, el resultado indica autocorrelación en los residuos.
* **Interpretación**: Dado que el p-valor es menor que 0.05, rechazamos la hipótesis de independencia, lo cual sugiere que los residuos están correlacionados. Esto puede indicar una limitación en el modelo, ya que los residuos no son completamente aleatorios.

**2. Supuesto de Homoscedasticidad**

* **Prueba de Breusch-Pagan**: El estadístico BP es **5.0707** con un p-valor de **0.02433**.
* **Interpretación**: El p-valor menor a 0.05 nos lleva a rechazar H0, indicando heteroscedasticidad. Esto implica que los residuos no tienen una varianza constante, lo cual podría afectar la robustez de las predicciones del modelo.

**3. Supuesto de Normalidad de los Residuos**

* **Prueba de Shapiro-Wilk**: Con W = **0.92146** y un p valor de **0.1781**, los residuos parecen seguir una distribución normal.
* **Interpretación**: Al no rechazar H0 (pues p es mayor que 0.05), concluimos que los residuos tienen una distribución normal, cumpliendo este supuesto clave y permitiendo inferencias válidas sobre el modelo.

h. Desarrolle un proceso de validación del modelo. Pronostique las ventas, construya un intervalo de confianza y de predicción individual del 95% para las ventas de licor local basado en las importaciones de un grupo de semanas registrados en la siguiente tabla. Interprete adecuadamente los resultados.

|  |  |
| --- | --- |
| Ventas | Importaciones |
| 495.2 | 120 |
| 360.0 | 175 |
| 112.8 | 300 |
| 102.5 | 360 |

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**Pronósticos e Intervalos Interpretados**:

1. **Para un nivel de importación de 120**:
   * **Estimación de ventas**: 482.92 millones de pesos
   * **Intervalo de confianza del 95%**: entre 403.17 y 562.68 millones de pesos, lo que refleja la expectativa del promedio de ventas en este nivel de importación.
   * **Intervalo de predicción del 95%**: entre 324.58 y 641.26 millones de pesos, mostrando el rango de posibles valores individuales de ventas, dado el nivel de importación.
2. **Para un nivel de importación de 175**:
   * **Estimación de ventas**: 378.71 millones de pesos
   * **Intervalo de confianza del 95%**: de 324.32 a 433.10 millones de pesos, representando el rango estimado para el promedio de ventas en este escenario.
   * **Intervalo de predicción del 95%**: entre 231.50 y 525.92 millones de pesos, reflejando la posible variabilidad de las ventas en semanas con este nivel de importaciones.
3. **Para un nivel de importación de 300**:
   * **Estimación de ventas**: 141.87 millones de pesos
   * **Intervalo de confianza del 95%**: entre 99.32 y 184.43 millones, sugiriendo que el promedio de ventas cae a medida que aumentan las importaciones.
   * **Intervalo de predicción del 95%**: de -1.38 a 285.13 millones de pesos, lo que implica una gran dispersión y variabilidad en las ventas individuales, que incluso podría llegar a valores muy bajos.
4. **Para un nivel de importación de 360**:
   * **Estimación de ventas**: 28.19 millones de pesos
   * **Intervalo de confianza del 95%**: entre -38.95 y 95.33 millones, lo que indica una expectativa de ventas bajas en promedio, con posibilidades de una reducción notable.
   * **Intervalo de predicción del 95%**: de -124.19 a 180.57 millones, reflejando una alta incertidumbre en ventas individuales en este escenario de alta importación.

**Interpretación General**:

* Los **intervalos de confianza** nos dan un rango estimado para el promedio de ventas esperado en función del nivel de importaciones, y muestran claramente que el promedio de ventas disminuye a medida que las importaciones aumentan.
* Los **intervalos de predicción**, considerablemente más amplios, capturan la variabilidad individual de las ventas. Para importaciones elevadas (300 y 360), los valores negativos en el intervalo sugieren que el aumento de productos importados puede tener un impacto sustancialmente negativo en las ventas locales, lo que indica una fuerte competencia en el mercado debido a la presencia de productos importados.

1. De acuerdo con los resultados obtenidos, considera que el modelo es adecuado para predecir las ventas de licores locales? Justifique adecuadamente su respuesta.

**Evaluación de la Adecuación del Modelo**: El modelo, respaldado por un coeficiente de determinación alto, describe adecuadamente la relación general entre las importaciones y las ventas locales de licor. Sin embargo, la presencia de autocorrelación y heteroscedasticidad en los residuos compromete la confiabilidad para pronósticos específicos. Los amplios intervalos de predicción, especialmente para niveles elevados de importación, indican una alta variabilidad en las ventas individuales, lo que sugiere limitaciones en la precisión para casos particulares.

**Conclusión**: Este modelo es util para observar tendencias generales entre importaciones y ventas, pero su efectividad disminuye al aplicarse a predicciones individuales. La dependencia de los residuos y la variabilidad de la varianza requieren atención. En consecuencia, el modelo es más adecuado para análisis globales, y podría necesitar ajustes adicionales para mejorar la precisión en estimaciones puntuales.