

Examen segundo parcial

⚠ Esta es una vista previa de la versión publicada del examen

Comenzado: 7 de sep en 10:19

Instrucciones del examen

El siguiente examen consta de dos temas.

Se debe adjuntar al final un Archivo.R, .rmd, archivo generado .html o archivo .txt . El archivo debe contener las respuestas a las preguntas.

Cuentan con 120 minutos. Distribuya bien su tiempo.

Es un examen a libro abierto, pueden utilizar el material de clase, páginas de internet, etc.

Es un examen individual, está prohibida la comunicación con sus compañeros o cualquier tercera persona durante el examen.

Cualquier acto de deshonestidad será informado a las autoridades competentes.

Al comenzar el examen, el estudiante acepta el siguiente compromiso:

COMPROMISO DE HONOR

Reconozco que el presente examen está diseñado para ser resuelto de manera individual, y no se permite la ayuda de fuentes no autorizadas ni copiar.

TEMA 1 (75 PTS)

La base de datos [puertos.csv](#) contiene datos de un estudio sobre los factores de productividad de las terminales de contenedores chinas.

Las variables que se incluyen son:

Puerto: Nombre del puerto

PuertoID: Identificación del puerto

Rendimiento: Rendimiento anual en millones de \$ por muelle

Area: Area en Km² por anclaje

Gruas_muelle: Número de grúas promedio por muelle

Gruas_terminal: Número de grúas promedio por terminal

Tractores: Tractores promedio por muelle

Longitud: longitud promedio por muelle

Profundidad: Profundidad del agua

Queremos modelar el rendimiento anual por muelle.

- a. Estime un modelo lineal aditivo de primer orden para la variable de respuesta usando todas las observaciones. **(5 pts)**
- b. Compruebe formalmente que el modelo no sufra de multicolinealidad. **(5 pts)**
- c. Grafique los residuos del modelo vs la interacción Area:Gruas_muelle. En base a lo que observa en el gráfico, ¿serviría incluir esta interacción en el modelo para reducir la variabilidad de los residuos? Sustente detalladamente su respuesta. **(5 pts)**
- d. Realice una búsqueda automática basado en BIC en ambas direcciones (stepwise regression), donde el modelo más completo sería modelo que incluya todas las variables explicativas en primer orden más la interacción Area:Gruas_muelle y el modelo más pequeño sólo el intercepto. **(10 pts)**
- e. Transforme la variable Área en una variable dicotómica que tome el valor de 1 si el área es mayor o igual a 200 (terminal grande) y 0 caso contrario (terminal pequeño). Reemplace esa variable por Área en el modelo seleccionado en el literal c y

estime un nuevo modelo. **(5 pts)**

f. Usando el modelo del literal e, ¿cuánto se estima en el modelo que sea el cambio promedio del rendimiento por un aumento de una grúa en el promedio de grúas por muelle si el terminal es grande ($Area_dicotómica=1$)? **(10 pts)**

g. Realice una prueba formal de hipótesis para comprobar que el aumento estimado en el literal e es significativo y concluya a un 5% de significancia. **(10 pts)**

h. Calcule el Mallow's Cp con su propia fórmula para los 3 modelos estimados (modelos de los literales a, d y e). ¿Existe algún modelo según Mallow's Cp que sea sesgado? Escoja el "mejor" modelo según este criterio. **(10 pts)**

i. En el modelo seleccionado en el literal h, obtenga intervalos de 99% confianza válidos para los coeficientes de las variables seleccionadas. Interprete los intervalos. **(10 pts)**

j. Existe alguna observación influyente en el modelo seleccionado en el literal h. De ser el caso, ¿en qué podría afectar a sus intervalos de confianza del literal i?

TEMA 2 (25 PTS)

Los siguientes resultados son de un modelo ANOVA de un factor (una vía) para el rendimiento anual por muelle de terminales japoneses según el tamaño del terminal (AREA2: pequeño, mediano, grande).

La codificación que se usó para el factor es la de efectos.

```

> aggregate(df2$Rendimiento~df2$Area2, FUN=mean)
  df2$Area2 df2$Rendimiento
1   Grande      659.0900
2  Mediano      641.4800
3  Pequeño      690.4933
> contrasts(df2$Area2)
      [,1] [,2]
Grande    1    0
Mediano   0    1
Pequeño  -1   -1
> summary(mod1)
          Df Sum Sq Mean Sq F value Pr(>F)
Area2      2   4809    2405   0.056  0.946
Residuals 10 431188    43119

```

Responda a las siguientes preguntas:

- ¿Cuántos terminales japoneses en total se tomaron en cuenta para este análisis? **(5 pts)**
- Suponga que quiere reportar estos resultados como un modelo de regresión lineal. Escriba la ecuación estimada del modelo de regresión lineal usando el contraste de efectos o suma. **(10 pts)**
- Escriba el contraste de hipótesis de la prueba F general del modelo de regresión lineal. Escriba el estadístico calculado, los grados de libertad y concluya a un 1% de significancia. **(10 pts)**

Pregunta 1

100 pts

Aquí debe subir el archivo con las respuestas a los temas 1 y 2

Cargar

Elegir un archivo

No guardado

Entregar examen