

## EXAMEN MÓDULO 3 – AÑO 2021

### Instrucciones

- Deben realizar los ejercicios en RMarkdown, en el que además del código R con todos los procesos que hayan realizado para ejecutar los ejercicios, deben incluir los comentarios y explicaciones del resultado de los mismos, cuestión que como se les ha ido indicando durante todo el trimestre, es lo más importante, que sepan interpretar los resultados.
- La resolución de sus casos deberán entregarla a través de la aplicación del curso virtual en el enlace del menú izquierdo “TAREAS” (conforme a las instrucciones dadas en el foro). **Tengan en cuenta que deben subir a la aplicación el fichero html que generen, NO el fichero de R.**
- La fecha máxima de entrega del trabajo es el 20 de marzo. Pueden subirlo antes si lo tienen terminado.

### Ejercicio 1

Instale la librería vars y utilice la base de datos Canada, que incluye las variables prod, e, U, rw. Buscando la ayuda sobre el fichero, indique que significa y como están medidas las cuatro variables.

Tenga presente que los datos están en formato de series temporales (ts). Si no desea trabajar en este formato, conviértalo a un data frame. ¿Qué periodo temporal cubren estos datos?

Realice un análisis descriptivo de las variables.

### Ejercicio 2

Estime un modelo MCO que relacione la productividad con los salarios. Responda:

- 1) ¿Son significativos los parámetros a un  $\alpha = 0,05$ ? Justifique su respuesta.
- 2) Obtenga un intervalo de confianza para la pendiente para un nivel de significación del 90%.
- 3) Evalúe la normalidad de los residuos a través de un histograma, un qq-plot y un test de normalidad. ¿Son normales los residuos? Justifique su respuesta.
- 4) Siendo datos de series temporales, verifique a través del contraste de Durbin Watson si existen problemas de autocorrelación.
- 5) Realice un gráfico que muestre los resultados de la regresión lineal simple. ¿Utilizaría este modelo para obtener pronósticos sobre la productividad?

### Ejercicio 3

Realice una regresión múltiple para explicar los salarios con todas las variables del data set.

- 1) Señale cuáles de las variables explicativas son significativas y cuales no lo son al 99% y al 95%. Justifique su respuesta.
- 2) Explique qué efecto tiene el desempleo en los salarios.
- 3) ¿Considera que los residuos de este modelo son aleatorios? Justifique su respuesta.
- 4) Presente un gráfico con los valores reales y los estimados, y evalúe los resultados obtenidos.
- 5) Cree usted que hay problemas de multicolinealidad entre los regresores.
- 6) Busque de forma automática el mejor modelo para los salarios de la industria manufacturera utilizando el método forward. Presente los sumarios y gráficos de los resultados, y justifique cual es el modelo que utilizaría para la estimación de los salarios.

#### **Ejercicio 4**

Con la sintaxis `rep(1:4, 84/4)` cree una variable cualitativa, que identifique cada trimestre del año. Evalúe la capacidad explicativa del factor trimestre en el empleo.

#### **Ejercicio 5**

Con las variables elegidas en el punto anterior, elija el modelo glm más adecuado utilizando las familias Gamma y Gaussiana. Obtenga para los dos modelos: el AIC, el BIC, el estadístico G2 de Wilks, la Chi Cuadrado de Pearson, la prueba de Hosmer-Lemeshaw, el  $R^2$  y el pseudo  $R^2$  de McFadden, pseudo  $R^2$  de Cox-Snell y pseudo  $R^2$  de Nagelkerke. A la vista de estos estadísticos responda razonadamente cuál de los dos modelos tiene mayor significación. Justifique su elección, y evalúe sus resultados.

#### **Ejercicio 6**

Si considera que hay problemas de colinealidad entre los regresores, plantee un modelo que los solucione.

#### **Ejercicio 7**

Dado que los datos que estamos analizando son indicadores macroeconómicos, en el análisis de datos outlier no se ha considerado la existencia de errores en los datos y tampoco se contempla la posibilidad de prescindir de ninguna observación de cara al análisis. Evalúe si existen o no problemas de observaciones demasiado influyentes en el análisis y establezca la conveniencia o no conveniencia de utilizar técnicas robustas.

#### **Ejercicio 8**

En el ámbito de la psicología diferencial, se ha diseñado un experimento con el objetivo de analizar si existen diferencias en la memoria a corto plazo en función del sexo. Para ello, se seleccionaron al azar 3 mujeres y 3 hombres de la misma edad, subdividiéndolos en 2 grupos

experimentales distintos. A cada uno de ellos, se le presentan, sucesivamente y de forma aleatoria, diferentes listas de palabras, registrándose, tras cada presentación, la cantidad de sustantivos y de adjetivos de la lista que recuerda cada sujeto.

La siguiente tabla muestra la matriz de datos del experimento, recogiendo la cantidad de palabras recordadas por el sujeto en función de su sexo y de la categoría gramatical a la que pertenece la palabra.

Sexo del sujeto	Sujetos	Sustantivos	Adjetivos
<b>Mujer</b>	S <sub>1</sub>	9	11
	S <sub>2</sub>	4	4
	S <sub>3</sub>	6	10
<b>Hombre</b>	S <sub>1</sub>	12	7
	S <sub>2</sub>	14	12
	S <sub>3</sub>	9	14

- 1) Con los datos que figuran en la tabla genere un data frame almacenando la información obtenida en el experimento.
- 2) Realice un análisis descriptivo del data frame generado, mediante tablas y gráficos.
- 3) Describa el diseño experimental utilizado, identificando los distintos tipos de variables independientes, y estime el modelo pertinente.
- 4) En vista de los resultados obtenidos en la descomposición de la variabilidad explicada y no explicada y de los correspondientes tests estadísticos, describa las conclusiones pertinentes.