



Titulación: **Ingeniero en Organización Industrial**

Asignatura: **Estadística Industrial**

Curso: **2006/2007**

EXAMEN FINAL JULIO 2007

Problema 1 (3 puntos)

Se piensa que la potencia eléctrica consumida al mes por una planta química está relacionada con la temperatura ambiente promedio ($^{\circ}\text{F}$), la pureza promedio del producto (%) y las toneladas de producto producidas. Los datos correspondientes a un año concreto se encuentran en el fichero **PRExJ07.sav**:

1. Realiza un análisis descriptivo previo de las variables del problema y comenta los resultados más relevantes. ¿Podemos suponer que nuestra variable respuesta es Normal? **(0.4 pts)**
2. Calcula la matriz de correlaciones de las cuatro variables. ¿Qué información proporciona esta matriz?. ¿Qué regresores del modelo presentan una más estrecha relación lineal entre sí?. ¿Cuál es la primera variable que debería entrar en el modelo? **(0.4 pts)**
3. Realiza la selección del modelo mediante regresión por pasos, hacia delante y hacia atrás. Comenta los resultados obtenidos. **(0.4 pts)**
4. ¿Qué modelo de regresión propondrías y por qué? Determina el modelo ajustado y comenta el valor de R^2 . **(0.4 pts)**
5. ¿Crees que podemos descartar un modelo de la siguiente forma? (Justifica tu respuesta) **(0.4 pts)**

$$\text{POTENCIA}_{\text{estimada}} = 200 + 1.2 * \text{TEMPERATURA}$$

6. Para el modelo resultante del apartado (4), guarda los residuos y los valores ajustados y estudia si se verifican las hipótesis del modelo de regresión múltiple, comentando los procesos utilizados. ¿Existen observaciones influyentes? **(0.6 pts)**
7. Proporciona una estimación puntual de la potencia que consumiría la planta química para un mes con temperatura ambiente promedio de 52°F , un intervalo de confianza para la potencia promedio consumida en ese caso, así como un intervalo de predicción. ¿Podemos asegurar que la potencia consumida no será superior a 300? ¿Y en promedio para los meses de esas características? **(0.4 pts)**

Problema 2 (2 puntos)

En el fichero **PCExJ07.sav** se encuentran los datos referidos al porcentaje de proteínas consumido a través de distintos alimentos (carnes rojas, carnes blancas, huevos, ..., frutas y verduras) correspondientes a 25 países europeos.

1. Realiza una clasificación de los países usando al menos **dos** métodos jerárquicos, justificando si conviene trabajar con los datos originales o tipificados. Comenta la medida de distancia utilizada en cada caso así como el tipo de enlace. **(0.5 pts)**
2. En función de los dendogramas obtenidos en los distintos métodos jerárquicos, indica cómo agruparías a los 25 países, es decir, determina el número de grupos que harías y qué países estarían en cada grupo. Intenta explicar qué caracteriza a los países de cada grupo. **(0.75 pts)**
3. Clasifica los países en 2 grupos usando el algoritmo de las k-medias. Determina si es más adecuado realizar 3 grupos en lugar de 2 estudiando cómo varían las sumas de cuadrados dentro de grupos (SCDG). **(0.75 pts)**

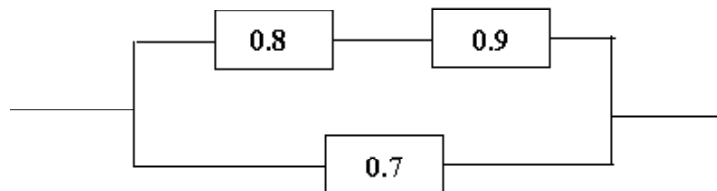
Problema 3 (4 puntos)

Usando los datos del fichero **Series_Julio07.sav** resuelve las siguientes cuestiones:

1. Representa la serie de datos en un gráfico temporal. ¿Crees que la serie proviene de un proceso estocástico estacionario? Justifica tu respuesta. **(0.25 pts)**
2. En el caso de que la serie no sea estacionaria, realiza las transformaciones que estimes oportunas para convertirla en estacionaria, comentando por qué las realizas. **(0.5 pts)**
3. Obtén el autocorrelograma simple y parcial de la serie estacionaria. En función de los resultados, ¿qué modelo(s) ARIMA propondrías como generador de la serie en estudio? Justifica tu respuesta. (Basta con que propongamos 2 modelos posibles). **(0.5 pts)**
4. Para cada uno de los modelos teóricos propuestos en el apartado anterior, responde a las siguientes cuestiones:
 - (a) Determina los coeficientes del modelo, justificando si son o no significativos. **(0.75 pts)**
 - (b) Determina valores indicativos de la bondad del ajuste. **(0.5 pts)**
 - (c) ¿El modelo es válido? Comprueba que se verifican las hipótesis sobre los residuos. **(0.75 pts)**
5. En función de los resultados del apartado anterior, comenta qué modelo o modelos te parecen adecuados para representar la serie en estudio. **(0.25 pts)**.
6. Realiza una predicción de la serie en estudio para los 20 instantes de tiempo siguientes (indica sólo las predicciones de los 3 primeros instantes). ¿Las 20 predicciones contienen el mismo error? Justifica tu respuesta. **(0.5 pts)**

Problema 4 (1 punto)

Se dispone de un sistema formado por tres componentes independientes dispuestas de la siguiente forma:



las cuales tienen una fiabilidad de cumplir su función durante 100 horas del 80%, 90% y 70%, respectivamente.

1. Determina la fiabilidad que tendrá el sistema de cumplir su función durante 100 horas. Suponiendo que cada componente puede ocupar cualquiera de las tres posiciones, ¿cómo colocarías las tres componentes de manera que se maximice la fiabilidad del sistema resultante? **(0.5 pts)**
2. Consideremos de nuevo el sistema inicial. Con el fin de aumentar la fiabilidad del sistema resultante, nos proponemos dos tipos de redundancia: redundancia a nivel de sistema y redundancia a nivel de componentes. ¿Qué tipo de redundancia aumentará más la fiabilidad? Calcula la fiabilidad de la opción que has considerado más adecuada. **(0.5 pts)**