



Dpto. Matemática Aplicada y Estadística

Titulación: **Ingeniero en Organización Industrial**

Asignatura: **Estadística Industrial**

Curso: **2009/2010**

**EXAMEN FINAL JUNIO 2010**

### Problema 1 (3.5 puntos)

En el fichero **PRegJun2010.sav** se encuentran datos relativos a partidos de la liga ACB de baloncesto. Se desea encontrar un modelo que explique los puntos por partido (**PUNTXPART**) que es capaz de anotar un jugador de baloncesto respecto a las siguientes variables regresoras: la altura del jugador (**ALTURA**), los minutos que juega por partido (**MINXPART**), los balones que pierde por partido (**BPXPART**), las faltas personales cometidas por partido (**FPXPART**) y el porcentaje en tiros de campo (**POR \_ TC**).

1. Las variables del problema, ¿presentan datos atípicos? ¿Podemos suponer que nuestra variable respuesta es Normal? Justifica tus respuestas **(0.3 ptos)**
2. En el caso de que la variable respuesta no provenga de una distribución Normal, realiza una transformación adecuada y explica por qué has decidido dicha transformación. **(0.4 ptos)**
3. Calcula la matriz de correlaciones de las variables del problema. ¿Qué regresores del modelo presentan una más estrecha relación lineal entre sí? (indica el valor que toma el coeficiente de correlación de Pearson). ¿Cuál es la primera variable que debería entrar en el modelo? (indica el valor que toma el coeficiente de correlación de Pearson). **(0.4 ptos)**
4. Realiza la selección del modelo mediante regresión por pasos, hacia delante y hacia atrás. Indica, para cada uno de los tres métodos, el modelo teórico resultante. Estudia si los modelos obtenidos son reducibles (simplificables) y si presentan multicolinealidad. **(0.7 ptos)**
5. ¿Qué modelo de regresión propondrías y por qué? Determina el modelo ajustado que explica los puntos por partido y comenta el valor de  $R^2$ . **(0.4 ptos)**
6. Para el modelo resultante del apartado anterior, guarda los residuos y los valores ajustados y estudia si se verifican las hipótesis del modelo de regresión múltiple y si existen observaciones influyentes, comentando los procesos utilizados. **(0.8 ptos)**
7. Proporciona una estimación puntual de los puntos por partido que anotaría un jugador de altura 2.05m, con un porcentaje en tiros de campo de 0.55 y que ha jugado 18 minutos, que ha perdido 5 balones y ha cometido 3 faltas por partido. ¿Podemos asegurar que los puntos por partido de dicho jugador será superior a 4? ¿Y en promedio para los jugadores de esas características? **(0.5 ptos)**

### Problema 2 (3 puntos)

En el fichero **PFactorialEx.sav** se encuentran los datos correspondientes a la contaminación del aire en 80 ciudades de US para el año 1960. Las variables se definen a continuación. TMR: Tasa de mortalidad, SMEAN: Media aritmética de cantidades de sulfato, SMAX: Cantidades máximas de sulfato cada dos semanas, PMEAN: Media aritmética de partículas suspendidas, PMAX: Cantidad máxima de partículas suspendidas, PM2: Densidad poblacional por milla cuadrada, GE65: Porcentaje poblacional con 65 años o más, PERWH: Porcentaje de Blancos en la población, NONPOOR: Porcentaje de familias con ingresos sobre el nivel de pobreza, LPOP: Logaritmo (base 10) de la población.

Se pretende aplicar una técnica de análisis multivariante que permita explicar la estructura de dependencia entre las variables del problema a partir de unos pocos factores.

1. Explica razonadamente qué tipo de análisis parece más adecuado para este estudio. **(0.2 ptos)**
2. Determina si el análisis propuesto es adecuado en función de los datos muestrales. Indica qué procedimientos permiten responder a esta cuestión, explica en qué consisten y comenta los resultados obtenidos con SPSS. **(0.6 ptos)**
3. Lleva a cabo el análisis propuesto, respondiendo a las siguientes cuestiones y comentando los procedimientos utilizados en cada caso:
  - (a) Determina el número de factores a retener. **(0.4 ptos)**

- (b) Obtén la matriz de cargas factoriales e indica el Modelo de Análisis Factorial propuesto (**indica al menos la primera y última fila del modelo**) ¿Qué método has utilizado para obtener las cargas factoriales? Comenta las ventajas e inconvenientes de este método frente a otros métodos que también permiten calcular las cargas factoriales. **(0.8 ptos)**
- (c) ¿Cómo podemos facilitar la interpretación de los factores retenidos? Aplícalo e interpreta los factores resultantes. **(0.5 ptos)**
- (d) Comenta la bondad del ajuste. Para ello, compara las correlaciones observadas con las reproducidas y determina si te parece adecuado el modelo factorial obtenido. ¿Qué representan las comunalidades? ¿Te parecen adecuadas en nuestro modelo? **(0.5 ptos)**

**Problema 3 (3.5 puntos)**

Usando los datos del fichero **PSeriesJun2010.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 ptos)**
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. **(0.4 ptos)**
3. Obtén el autocorrelograma simple y parcial de la serie estacionaria. En función de los resultados, ¿qué modelo(s) ARIMA propondrías como generador(es) de la serie en estudio? Justifica tu respuesta. **(0.5 ptos)**
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, y proporciona el modelo irreducible correspondiente. **(0.5 ptos)**
  - (b) Determina valores indicativos de la bondad del ajuste. **(0.25 ptos)**
  - (c) ¿El modelo es válido? Comprueba que se verifican las hipótesis sobre los residuos. **(0.75 ptos)**
5. **(0.45 ptos)**. En función de los resultados del apartado anterior, indica qué modelo te parecen más adecuado para representar la serie en estudio y exprésalo en la forma:

$$X_t = cte + a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} - b_1\varepsilon_{t-1} - b_2\varepsilon_{t-2} - \dots - b_q\varepsilon_{t-q} + \varepsilon_t$$

donde  $X_t$  representa la serie original en estudio.

6. Realiza una predicción de la serie 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. **(0.4 ptos)**