



Dpto. Matemática Aplicada y Estadística

Titulación: **Ingeniero en Organización Industrial**

Asignatura: **Estadística Industrial**

Curso: **2008/2009**

EXAMEN FINAL ENERO 2009

Problema 1 (3.75 puntos)

El fichero **PRegEne09.sav**, contiene los datos correspondientes a la observación de 21 días de trabajo en una planta química para la oxidación del amonio como una etapa en la producción del ácido nítrico. Las variables observadas son X_1 (flujo de aire), X_2 (temperatura del agua de refrigeración en $^{\circ}C$), X_3 (concentración de ácido (%)) e Y (pérdida acumulada, porcentaje del amonio que escapa sin ser absorbido).

El objetivo del estudio es encontrar un modelo que explique la pérdida acumulada de amonio en función de las otras tres variables.

1. Indica la variable respuesta y los regresores del problema. Las variables del problema, ¿presentan datos atípicos? ¿Podemos suponer que nuestra variable respuesta es Normal? Justifica tus respuestas **(0.4 pts)**
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 pts)**
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 pts)**
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. **(0.75 pts)**
5. ¿Qué modelo de regresión propondrías y por qué? Determina el modelo ajustado y comenta el valor de R^2 . **(0.5 pts)**
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, comentando los procesos utilizados. **(0.5 pts)**
7. ¿Existen observaciones influyentes? Comenta los criterios utilizados. **(0.3 pts)**
8. Proporciona una estimación puntual de la pérdida de amonio cuando $X_1 = 60$, $X_2 = 20$, $X_3 = 60$. ¿Podemos asegurar que la pérdida de amonio será inferior a 1.8? ¿Y en promedio para los días de esas características? **(0.5 pts)**

Problema 2 (2.25 puntos)

En el fichero **transportes1997_Cluster.sav** se encuentran los datos económicos correspondientes a las empresas de transporte de viajeros (por carretera) para las distintas comunidades autónomas españolas y referidos al año 1997 (fuente: www.ine.es). Las variables en consideración son: gastos en personal, gastos en material consumible, gastos en servicios exteriores, amortizaciones, otros gastos de explotación e ingresos y subvenciones (unidades: miles de pesetas).

1. Realiza una clasificación de las comunidades autónomas 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 las 17 comunidades autónomas, es decir, determina el número de grupos que harías y qué comunidades estarían en cada grupo. Intenta explicar qué caracteriza a las comunidades de cada grupo. **(0.75 pts)**

3. Usando el algoritmo de las k-medias, determina cuál sería el número de grupos adecuado estudiando cómo varían las sumas de cuadrados dentro de grupos (SCDG). Una vez determinado el número de grupos a realizar, clasifica las comunidades autónomas usando el algoritmo de las k-medias e interpreta los resultados. **(1 pto)**

Problema 3 (4 puntos)

En el fichero *hoteles_ST.sav* se encuentran los datos correspondientes a la ocupación hotelera en la región de Murcia (número de personas), recopilados mensualmente desde enero de 2004 hasta noviembre de 2008 (fuente: www.ine.es).

1. Representa los datos de la ocupación hotelera en un gráfico temporal. ¿La serie presenta Estacionalidad? **(0.25 ptos)**
2. Justifica si se trata de un modelo aditivo o multiplicativo. Para confirmarlo, realiza un gráfico de desviaciones típicas frente a medias para cada año (representalo en tu folio a mano alzada). **(0.25 ptos)**
3. Extrae las componentes de la serie (Tendencia-Ciclo, Estacionalidad e Irregular). Indica los 3 primeros valores de cada componente. **(0.25 ptos)**
4. ¿Cómo dirías que es la tendencia? Intenta dar una explicación a la forma que presenta. **(0.25 ptos)**
5. Proporciona un modelo determinista que nos permita realizar predicciones, **justificando qué curva has seleccionado para representar la tendencia. (0.5 ptos)**. Expresa el modelo de la forma:

$$OCUPACION_HOTELERA_ESTIMADA = \dots$$

6. Determina los residuos del modelo anterior, es decir, la diferencia entre el valor observado y el estimado por el modelo determinista. Indica los 3 primeros residuos. **(0.25 ptos)**
7. Determina las predicciones de la ocupación hotelera para el próximo año 2009. (Indica sólo las predicciones para los meses de enero, agosto y diciembre). **(0.5 ptos)**

Nos proponemos ahora analizar la serie del fichero *hoteles_ST.sav* mediante una técnica de alisado exponencial.

8. ¿Qué técnica de alisado exponencial te parece más adecuada para analizar esta serie? Razona tu respuesta. **(0.25 ptos)**
9. Aplica la técnica de alisado que has considerado adecuada, seleccionando como parámetros de alisado aquellos que minimizan la suma de cuadrados de los errores (búsqueda en rejilla con precisión de 0.1). Determina cómo quedarían las fórmulas recurrentes de las series alisadas e interpreta el significado de los parámetros de alisado obtenidos. **(0.75 ptos)**
10. Compara los residuos obtenidos mediante esta técnica de alisado con los residuos de la descomposición clásica. Para ello, realiza un gráfico comparativo de los residuos: representalo en tu folio a mano alzada y coméntalo. **(0.25 ptos)**
11. Determina las predicciones de la ocupación hotelera para el próximo año 2009. (Indica sólo las predicciones para los meses de enero, agosto y diciembre). **(0.25 ptos)**
12. Si queremos que nuestro método de alisado de más importancia al cambio de tendencia que se produce al final de la serie, ¿cómo deberíamos seleccionar los parámetros de alisado?. Justifica tu respuesta. **(0.25 ptos)**