



Dpto. Matemática Aplicada y Estadística

Titulación: **Ingeniero en Organización Industrial**

Asignatura: **Estadística Industrial**

Curso: **2006/2007**

EXAMEN FINAL ENERO 2007

Problema 1 (5 puntos)

El fichero **PRegresionEx.sav**, contiene los datos correspondientes a la presencia (en %) de cinco componentes químicos en un tipo de cemento, así como el calor emitido (en calorías por gramo de cemento) durante el proceso de endurecimiento. Se desea proponer un modelo que permita predecir el calor emitido en función de los componentes químicos presentes en el cemento.

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 **(0.5 pts)**
2. Calcula la matriz de correlaciones de las seis variables. ¿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). **(0.5 pts)**
3. 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. **(1 pts)**
4. Teniendo en cuenta el apartado anterior, determina el modelo(s) ajustado(s) que consideres más adecuado(s) y comenta el valor del R^2 (coeficiente de determinación) **(0.25 pts)**
5. Estudia si el modelo que contiene a los regresores A y B presenta observaciones influyentes (comenta los criterios utilizados). Realiza un análisis de los residuos para estudiar si se verifican las hipótesis del modelo de regresión múltiple (comenta brevemente los procesos utilizados). **(0.75 pts)**
6. Obtén una estimación puntual del calor emitido por el cemento sabiendo que $A = 15$ y $B = 30$. Determinar también un intervalo de confianza para el calor emitido en ese caso, así como un intervalo de predicción. ¿Podemos asegurar que el calor emitido por el cemento superará las 95 cal/gr? ¿Y en promedio? **(0.5 pts)**

Por otra parte se pretende aplicar una técnica de análisis multivariante que permita reducir la dimensión de los datos correspondientes a la presencia de los componentes químicos, perdiendo muy poca información al reducir la dimensión. Obsérvese que los datos de los que partimos son 5-dimensionales, pues para cada muestra de cemento se anotan los porcentajes de los cinco componentes químicos presentes.

7. Explica razonadamente qué tipo de análisis parece más adecuado para este estudio. **(0.25 pts)**
8. Para realizar el análisis propuesto, conviene trabajar con la matriz de correlaciones en lugar de la matriz de covarianzas. ¿Por qué? **(0.25 pts)**
9. ¿Con cuántas componentes principales te quedarías? Justifica tu respuesta comentando los distintos métodos de selección y los resultados obtenidos con SPSS. **(0.5 pts)**
10. Determina cómo se expresarían las componentes principales que has decidido retener, indicando además los valores y vectores propios de la matriz de correlaciones. **(0.5 pts)**

Problema 2 (4 puntos)

En el fichero **Series_Ene07.sav** se encuentran los datos correspondientes a la producción de CDs de una determinada empresa fabricante (en miles de unidades), recopilados mensualmente desde enero de 1995.

1. Representa los datos de la producción de CDs en un gráfico temporal. ¿La serie presenta Estacionalidad? ¿cómo dirías que es la tendencia? **(0.25 pts)**

- Determina si se trata de un modelo aditivo o multiplicativo. Para ello, realiza un gráfico de desviaciones típicas frente a medias para cada año (representalo en tu folio a mano alzada). **(0.25 ptos)**
- Extrae las componentes de la serie (Tendencia-Ciclo, Estacionalidad e Irregular). Indica los 3 primeros valores de cada componente. **(0.5 ptos)**
- Proporciona un modelo determinista que nos permita realizar predicciones, **considerando que la tendencia es lineal**. **(0.5 ptos)**. Expresa el modelo de la forma:

$$PRODUCCION_ESTIMADA = \dots$$

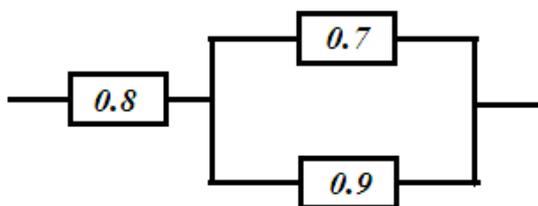
- Determina los residuos del modelo anterior, es decir, la diferencia entre el valor observado y el estimado por el modelo determinista. Indica los 3 primeros residuos. **(0.25 ptos)**
- Determina las predicciones de la producción de CDs para los cuatro primeros meses del año 2007. **(0.5 ptos)**

*Nos proponemos ahora analizar la serie del fichero **Series_Ene07.sav** mediante una técnica de alisado exponencial.*

- ¿Qué técnica de alisado exponencial te parece más adecuada para analizar esta serie? Razona tu respuesta. **(0.25 ptos)**
- 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)**
- 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. **(0.5 ptos)**
- Determina las predicciones de la producción de CDs para los cuatro primeros meses del año 2007. **(0.25 ptos)**

Problema 3 (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%, 70% y 90%, respectivamente.

- 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? **(0.3 ptos)**
 - 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. **(0.3 ptos)**
- Se sabe que la función de razón de fallo de un determinado componente electrónico viene dada por una curva de bañera. Representa gráficamente cómo sería la función de razón de fallo, explicando el significado de cada tramo. Indica además un modelo de distribución (así como el valor de sus parámetros) cuya razón de fallo permita representar cada tramo de la curva de bañera. **(0.4 ptos)**