EL USO DE COMPONENTES PRINCIPALES EN LA MONITORIZACION Y DIAGNOSTICO DE PROCESOS INDUSTRIALES

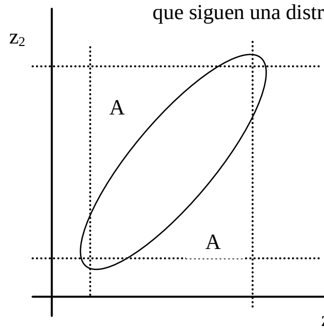
Pagura, José Alberto

Instituto de Investigaciones Teóricas y Aplicadas. Escuela de Estadística.

<u>Control estadístico de procesos tradicional</u>: analiza un pequeño conjunto de variables de calidad de forma individual y monitoriza el proceso, realizando a partir de las señales de fallo que se detecten, las acciones correctivas pertinentes.

<u>Control multivariado de procesos-enfoque tradicional</u>: estudia la distribución conjunta de las variables y, a partir de ella detecta señales de fallo.
Ventaja con respecto al control estadístico univariado: un fallo puede estar dado no solo por la existencia de un valor fuera de los límites de control, sino también por una desproporción en los valores de las variables en forma conjunta.
Si se estudian M variables que siguen una distribución normal multivariada, $N_M(\mu,\Sigma)$ resulta adecuado un límite superior de control dado o por la distribución χ^2, si la matriz de covariancias es conocida, o por la T^2 de Hotelling, si la matriz de covariancias se estima a partir de una muestra.

Ventajas de un diagrama de control multivariado. Caso de dos variables z_1 y z_2 que siguen una distribución conjunta normal



$\rho>0$
$\sigma_2 > \sigma_1$

Las líneas punteadas indican los límites de control para los diagramas univariados de x_1 y x_2. La zona A corresponde a fallos no detectados por los diagramas de control univariados pero si por el T^2

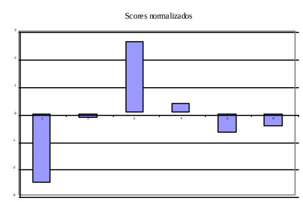
<u>Variables responsables de las señales de salidas de control</u>
Ante una señal de salida de control, se deben encontrar las variables que la provocaron, con el fin de realizar las acciones correctivas necesarias. ¿Como hacerlo de manera rápida, fácil y efectiva?

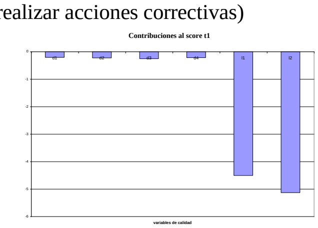
<u>Errores normalizados</u>: graficar, para cada variable en forma individual los valores $(z_{ij}-m_j)/s_j$. Es similar al control univariado para cada variable; *no se detectaran como anomalías, las observaciones en la zona A del gráfico anterior.*

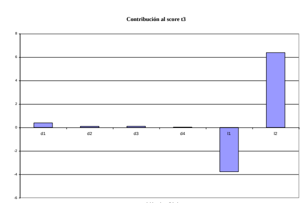
<u>Scores normalizados</u>: T^2 puede ser descompuesta a partir de los autovectores $\mathbf{p}_1,\mathbf{p}_2,...,\mathbf{p}_M$ y de los autovalores correspondientes $\lambda_1,\lambda_2,...,\lambda_M$ de la matriz de covariancias muestral S.

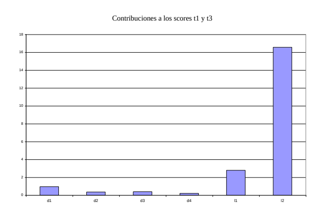
$T^2=\sum_{j=1}^{M}\frac{t_j^2}{\lambda_j}$ con scores $t_j = p_j^{'}(z-\mu)$ (descomposición en componentes principales)
Donde:
Z matriz cuyas filas son las unidades y las columnas corresponden a las variables que se estudian.
P: matriz cuyas columnas son los autovectores de la matriz **S** (cargas)
T: matriz que contiene en cada fila los scores correspondientes a una unidad
Entonces: **T**=**ZP** , también **Z**=**TP**'
Se eligen los scores mayores que un valor de referencia y se observan las contribuciones de cada variable a los estos scores.

<u>Ejemplo</u>: en un proceso de producción de prendedores, se controlan seis variables de calidad, cuatro diámetros y dos longitudes. Se obtuvieron $\hat{\mu}=m$ y $\hat{\Sigma}=S$.
Para la monitorización se fijó T^2=14,10. Ante una señal de salida de control, se calculan y grafican los scores normalizados que en este caso, muestran significativos al primero y tercero (con un α=0.05, por el método de Bonferroni, se toman como valores críticos ±2.63). Los gráficos de contribuciones muestran a la longitud 2 como variable sobre la cual realizar acciones correctivas)



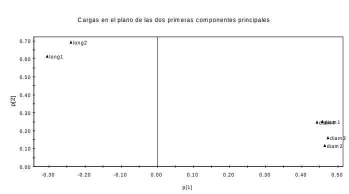


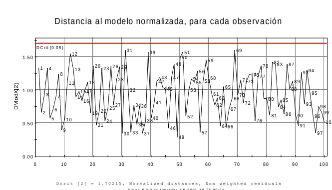


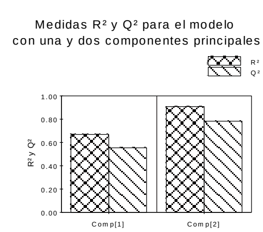


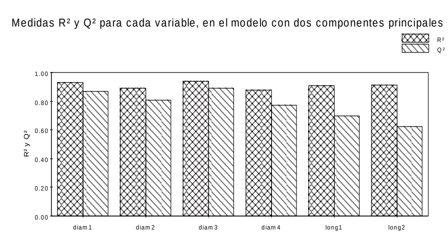
<u>Análisis en componentes principales</u>
La información contenida por la matriz **Z** podrá explicarse con un número de variables A mucho menor que M. Estas variables serán los A primeros scores o componentes principales $\mathbf{t}_1, \mathbf{t}_2,...,\mathbf{t}_A$.
Si las componentes principales fueron calculadas a partir de **S**, estarán afectadas por las diferentes escalas en que están medidas las variables z_j por lo que previamente las variables se escalan, en la mayoría de los casos dividiendo por su desviación estándar (hay otras formas de escalamiento).
Teniendo en cuenta que **Z**=**T*****P**', puede escribirse el modelo de componentes principales: **Z**=**T***$\mathbf{P'_A}$ +**E**, con $\mathbf{P'_A}$ matriz cuyas columnas son los primeros A vectores propios considerados y **E** matriz de residuales.

<u>Herramientas de diagnóstico</u>
Gráficos de scores en los planos definidos por las variables latentes junto a la elipse definida por la T^2 de Hotelling , para un nivel de significación dado. Una observación "fuera" de la elipse indicará una anomalía(outliers severos). Luego se podrá analizar que variable introdujo la anomalía.
Grafico de las cargas. Permite visualizar las asociaciones entre las variables. La representación conjunta con las observaciones permitirá encontrar explicaciones sobre qué variables causan los agrupamientos y tendencias que se observan.
Distancia DmodZ. Es la distancia de una observación, al modelo, calculada a partir de los residuales. Permite la detección de "ouliers moderados". Al normalizar esta distancia y elevarla al cuadrado, se obtiene una estadística que sigue una distribución $F_{g1,g2}$, que permitirá evaluar si la observación está o no bien explicada por el modelo.
Se calculan también las medidas $\mathbf{R^2}$ y $\mathbf{Q^2}$ para evaluar el porcentaje de variabilidad explicado por el modelo y la capacidad de predicción del mismo, datos útiles a la hora de elegir el número de componentes principales a incluir. Estos valores se calculan también, para cada variable, permitiendo evaluar la importancia de cada una de ellas en el modelo.









Los gráficos anteriores corresponden a un modelo con dos componentes principales (los vectores propios acumulan en 91%), considerado apto teniendo en cuenta los indicadores establecidos.

<u>La monitorización del proceso</u>.
Se registrarán datos en condiciones normales de operación, se analizarán decidiendo el número A de componentes principales a utilizar. Para cada nueva observación se calcularán el valor de T² a partir de los A scores y la distancia al modelo normalizada.