Pub. Mat. UAB
Nº 22 Nov. 1980
Actes VII JMHL

NOTAS SOBRE DISEÑOS ROBUSTOS USANDO g-INVERSA GENERALIZADA. APLICACION A DISEÑOS DE PESADAS (WEIGHING)

Herminia Inmaculada Calvete Fernández

Dpto. de Cálculo de Probabilidades y Estadística Matemática
Universidad de Zaragoza

A robust design is studied proposed by George Box and Norman Draper provided that the design matrix rank is non-maximum. In this case analogous results are shown. Finally an application to the weighing designs case is given.

1.- Diseños robustos usando g-inversa generalizada.

Sea el modelo lineal

$$Y = X\beta + \varepsilon$$

donde Y es un vector aleatorio n-dimensional de observaciones; X es una matriz $n\times p$ $(p \leq n)$ conocida llamada matriz del diseño; β es un vector p-dimensional de parámetros desconocidos cuyos valores pertenecen a R^p; y ε es un vector aleatorio n-dimensional de medias nulas y matriz de covarianzas $\sigma^2 I_n$. Supondremos además que X es de característica $s < p$

Aplicando la técnica de mínimos cuadrados, el estimador $\hat{\beta}$, de β, debe satisfacer la ecuación:

$$(X'X)\hat{\beta} = X'Y$$

Como $\operatorname{rang}(X'X) = \operatorname{rang}(X) = s < p$, no existe inversa. Introduciendo la inversa generalizada de Moore-Penrose (que existe y es única), la solución general de la ecuación será:

$$\hat{\beta} = (X'X)^+(X'Y) + [(X'X)^+(X'X) - I]Z$$

siendo Z una matriz arbitraria.

La estimación de $X\beta$ será:

$$\hat{Y} = X\hat{\beta} = X(X'X)^+(X'Y) + X[(X'X)^+(X'X) - I]Z = RY + K$$

La matriz R es idempotente y simétrica.

Supongamos que la u-ésima observación ha sufrido una alteración debida a la "aberración" c . El cambio que se produce en $\hat{y}_i$ vendra dado por $\zeta_i = c.r_{iu}$

Sea $\zeta = (\zeta_1 \ldots\ldots \zeta_n)'$. Se define la discrepancia global causada por el efecto de c sobre la u-ésima observación como:

$$d_u = \zeta' \cdot \zeta = c^2 \sum_{i=1}^{n} r_{iu}^2 = c^2 . r_{uu}$$

por la idempotencia de R.

Si c pudiera afectar a cualquiera de las n observaciones dando origen a $d_1; \ldots ; d_n$ la discrepancia media sería:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i = \frac{c^2}{n} \sum_{u=1}^{n} r_{uu}$$

$$\mathrm{traz}(R) = \mathrm{traz}\left[(X'X)^{+}(X'X)\right] = \mathrm{rang}\left[(X'X)^{+}(X'X)\right] = \mathrm{rang}(X'X) = s$$

$$\bar{d} = \frac{c^2 s}{n}$$

es decir, la discrepancia media inducida en $\hat{Y}$ es constante e independiente del diseño.

Si se desea elegir el diseño de forma que en ningún punto pueda, la adición de c, inducir una discrepancia excesivamente grande, habrán de hacerse las d_u tan uniformes como sea posible. Una medida conveniente de la uniformidad es:

$$V(d) = \frac{1}{n} \sum_{u=1}^{n} (d_u - \bar{d})^2 = \frac{c^4}{n}\left[\sum_{u=1}^{n} r_{uu}^2 - \frac{s^2}{n}\right] =$$

$$= \frac{c^4}{n}\left[r - \frac{s^2}{n}\right]$$

siendo $r = \sum_{u=1}^{n} r_{uu}^2$

Para asegurar la insensibilidad del diseño a las observaciones erróneas habrá de minimizarse V(d) lo que es equivalente a la minimización de r cuando n y s están fijados.

De los razonamientos anteriores se puede concluir que el criterio "r" dado por Box y Draper puede generalizarse con los mismos resultados al caso en que, por no poderse asegurar la existencia de matrices inversas, se empleen matrices inversas generalizadas.

2.- Aplicación del artículo de Box y Draper a los diseños de pesadas (weighing).

Supongamos que se quiere determinar los pesos de p objetos usando una balanza química (balanza de dos brazos) en la que se pueden hacer n pesadas.

En cada una de las pesadas se toman dos decisiones:

i) si un objeto dado se incluye o no en la pesada

ii) ¿ en qué platillo se coloca ?

Definimos x_{ij} $(i = 1,\ldots n;\ j = 1,\ldots,n)$

$x_{ij} = 1$ si el j-ésimo objeto está en el platillo izdo durante la i-ésima pesada.

$x_{ij} = -1$ si el j-ésimo objeto está en el platillo dcho durante la i-ésima pesada

$x_{ij} = 0$ si el j-ésimo objeto no se incluye en la i-ésima pesada

La matriz $X = (x_{ij})$ caracteriza completamente el experimento de pesadas.

Sean $W' = (w_1,\ldots,w_p)$ los verdaderos pesos de los p objetos e $Y' = (y_1,\ldots,y_n)$ los resultados de las n pesadas, es decir, las lecturas que proporciona la balanza. Estas lecturas pueden representarse po el modelo lineal:

$$Y = X\,W + \varepsilon$$

donde $\varepsilon' = (\varepsilon_1,\ldots,\varepsilon_n)$ y ε_i es el error entre la lectura observada y la esperada.

Supondremos que ε es un vector aleatorio con medias cero y matriz de covarianzas $\sigma^2 I$. Esta suposición es razonable en el caso de que los objetos tengan masa puequeña en comparación con la masa de la balanza.

Suponiendo X de rango p, la estimación, por mínimos cuadrados, de los verdaderos pesos es:

$$\hat{W} = (X'X)^{-1}X'Y$$

y su matriz de covarianzas es:

$$\sigma^2(X'X)^{-1}$$

Hotelling (1944) demostró que para cualquier diseño de pesadas la varianza de $\hat{w}_i$ $(1 \leq i \leq p)$ es mayor o igual que σ^2/n. Así, un diseño X se dirá óptimo si las estimaciones de cada peso tienen mínima varianza σ^2/n.

Puede verse (Raghavarao) que X es óptima si y sólo si $X'X = I_p.n$. Luego cualesquiera p ($\leq$ n) columnas de una matriz de Hadamard de orden n constituyen un diseño nxp de pesadas óptimo.

Consideremos un diseño de pesadas óptimo en el que la matriz X esté formada por p ($\leq$ n) columnas de una matriz de Hadamard.

En este caso

$$\hat{W} = \frac{1}{n} X'Y$$

y la estimación de Y será

$$\hat{Y} = X\hat{W} = \frac{1}{n} (XX')Y = R.Y$$

Suponemos que la u-ésima observación ha sufrido una alteración debida a la "aberración" c. Con la misma notación del apartado anterior en este caso se obtendrá:

$$d_u = c^2.r_{uu}$$

$$\bar{d} = \frac{c^2}{n} p$$

$$V(d) = \frac{c^4}{n} \left[r - \frac{p^2}{n} \right]$$

Los elementos de la diagonal principal de R valen todos p/n, por tanto

$$r = \sum_{u=1}^{n} r_{uu}^2 = \sum_{u=1}^{n} \frac{p^2}{n^2} = \frac{p^2}{n}$$

Luego esta matriz X proporciona el mínimo de la V(d), V(d) = 0, y por ello es óptima para el criterio "r".

3.- Bibliografía.

- Box, G.; Draper, N. "Robust designs". Biometrika (1975) 62,2
- Hotelling, H. "Some improvements in weighing and other experimrntal techniques" Ann. Math. Statist. 15, 297-306. (1944)
- Raghavarao, D. "Constructions and combinatorial problems in design of experiments". Wiley (1971)