Revista Colombiana de Estadística
No.31 y No. 32

# ANÁLISIS DEL DESVÍO : UNA ALTERNATIVA AL ANÁLISIS DE VARIANZA

MIGUEL A. DUMETT

Profesor Asistente
Universidad Nacional de Colombia

ABSTRACT. Se muestra como el análisis de desvío de los modelos lineales generalizados es una alternativa al análisis de varianza.

## Introducción

Una de las metodologías estadísticas más utilizadas en Ciencias Agropecuarias es sin duda el análisis de varianza. Este tipo de estudio descompone la suma de cuadrados total en la suma de cuadrados de los efectos que inciden en la variable respuesta. Esto permite, en caso de tenerse un modelo balanceado con efectos fijos, la posibilidad de responder de manera práctica a cuestiones tales como si un tratamiento es mejor o no que otro, (por medio de comparaciones planeadas de efectos de tratamientos) u otras semejantes por medio de otras pruebas.

Debido al éxito alcanzado por los modelos estadísticos basados en el análisis de varianza, en la práctica se suele asumir como verdaderos los supuestos de éste. Uno de ellos, es la distribución normal de los errores; suposición de hecho muy fuerte pero intuitivamente razonable y que lo único que aporta a la metodología es la formación de

*Key words and phrases.* Análisis de varianza, Desvío, modelos lineales generalizados, GLIM.

intervalos de confianza para los parámetros involucrados en el modelo y la posibilidad de docimar hipótesis acerca de estos.

Sin embargo, estas herramientas de estimación son esenciales en la práctica y por ello no puede omitirse una suposición distribucional para los errores. Pero, lo que si puede hacerse es cambiar la distribución por otra y tratar de construir un análogo al análisis de varianza. Surge entonces el problema de cual debe ser esa distribución alternativa.

Una de las motivaciones de realizar este trabajo es mostrar como los modelos lineales generalizados son una alternativa válida para proponer modelos estadísticos a diferentes situaciones. En ellos, existen varios aspectos relevantes, uno de los cuales es la versatilidad para la elección de la distribución del error y otro el análisis de desvío (que consiste en descomponer la función desvío del modelo en la suma de desvíos que inciden en la variable respuesta). Este análisis es una herramienta muy poderosa pero suele ser desconocida en nuestro medio. Pretendemos efectuar una aproximación a este análisis a partir del análisis de varianza con efectos fijos balanceados aplicado a un caso particular.

Siguiendo los lineamientos usuales en los modelos lineales generalizados descritos por Jorgensen (1992), se utilizará la familia de distribuciones denominada modelos de dispersión exponencial y que incluye a la familia exponencial natural como una subfamilia.

A continuación indicamos el orden que se seguirá en este trabajo. Primero se presenta como trabaja el análisis de varianza para un modelo de de efectos fijos cruzados balanceados con 4 factores; luego, se introduce la familia de modelos de dispersión, entre cuyos miembros se encuentra la distribución normal (esencial en el análisis de varianza) y se indica que el sentido de la generalización será permitir otras distribuciones diferentes a la normal con media cero y varianza constante para los errores. A continuación, se introduce la función desvío, la cual en el caso de suponer una distribución normal coincidirá con la suma de cuadrados clásica. Finalmente, se indica en que consiste el análisis del desvío; como la metodología correspondiente es análoga con la que se realiza en el análisis de varianza y se ilustra esto con dos ejemplos.

## 1. Modelo de Efectos Fijos Cruzados Balanceados

Para ser más explícitos supongamos un diseño de 4 factores. Como lo muestran Johnson y Leone (1964), el modelo estadístico correspondiente se escribe como

$$\begin{aligned} y_{tijkl} = A + B_t + C_i + D_j + E_k + (BC)_{ti} + (BD)_{tj} \\ + (BE)_{tk} + (CD)_{ij} + (CE)_{ik} + (DE)_{jk} + (BCD)_{tij} + (BCE)_{tik} \\ + (BDE)_{tjk} + (CDE)_{ijk} + (BCDE)_{tijk} + \epsilon_{tijkl} \end{aligned} \tag{1.1}$$

$$t = 1, \cdots, b; \quad i = 1, \cdots, c; \quad j = 1, \cdots, d; \quad k = 1, \cdots, e; \quad l = 1, \cdots, n$$

junto a

$$\sum_t B_t = 0; \quad \sum_i C_i = 0; \quad \sum_j D_j = 0; \quad \sum_k E_k = 0; \tag{1.2}$$

$$\begin{array}{llll} \sum_t (BC)_{ti} = 0; & \sum_i (BC)_{ti} = 0; & \sum_t (BD)_{tj} = 0; & \sum_j (BD)_{tj} = 0; \\ \sum_t (BE)_{tk} = 0; & \sum_k (BE)_{tk} = 0; & \sum_i (CD)_{ij} = 0; & \sum_j (CD)_{ij} = 0; \\ \sum_i (CE)_{ik} = 0; & \sum_k (CE)_{ik} = 0; & \sum_j (DE)_{jk} = 0; & \sum_k (DE)_{jk} = 0; \end{array}$$

$$\sum_t (BCD)_{tij} = 0; \quad \sum_i (BCD)_{tij} = 0; \quad \sum_j (BCD)_{tij} = 0;$$
$$\sum_t (BCE)_{tik} = 0; \quad \sum_i (BCE)_{tik} = 0; \quad \sum_k (BCE)_{tik} = 0;$$
$$\sum_t (BDE)_{tjk} = 0; \quad \sum_j (BDE)_{tjk} = 0; \quad \sum_k (BED)_{tjk} = 0;$$
$$\sum_i (CDE)_{ijk} = 0; \quad \sum_j (CDE)_{ijk} = 0; \quad \sum_k (CDE)_{ijk} = 0;$$
$$\sum_t (BCDE)_{tijk} = 0; \qquad \sum_i (BCDE)_{tijk} = 0;$$
$$\sum_j (BCDE)_{tijk} = 0; \qquad \sum_k (BCDE)_{tijk} = 0$$

Se debe entender que cada una de las sumas anteriores debe ser igual a cero para cualesquiera de los valores que tomen los subíndices que aparecen sobre los cuales no se efectúa dicha suma. Además, aunque esta forma de imponer restricciones no estimables sobre los efectos de los tratamientos no es la usual, las estándares se siguen de éstas. Por ejemplo, usualmente se escribe

$$\sum_t (BC)_{ti} = \sum_i (BC)_{ti} = \sum_{ti} (BC)_{ti} = 0$$

lo cual es redundante, ya que de los dos primeros sumandos se sigue el tercero pues

$$\sum_{ti} (BC)_{ti} = \sum_t \Big( \sum_i (BC)_{ti} \Big) = 0.$$

También se consideran los siguientes supuestos

$$E(\epsilon_{tijkl}) = 0 \tag{1.3a}$$

$$Var(\epsilon_{tijkl}) = \sigma^2 \tag{1.3b}$$

$$Cov(\epsilon) = \sigma^2\, I \tag{1.3c}$$

$$\epsilon_{tijkl} \sim N(0, \sigma^2) \tag{1.3d}$$

En estos modelos pueden construirse tablas de análisis de varianza para probar las hipótesis más usuales entre las cuales están las siguientes

$$H_o : B_1 = \cdots = B_b = 0 \tag{1.4}$$

$$H_o : C_1 = \cdots = C_c = 0$$

$$H_o : D_1 = \cdots = D_d = 0$$

$$H_o : E_1 = \cdots = E_e = 0$$

$$H_o : (BC)_{11} = \cdots = (BC)_{bc} = 0$$

$$H_o : (BD)_{11} = \cdots = (BD)_{bd} = 0$$

$$H_o : (BE)_{11} = \cdots = (BE)_{be} = 0$$

$$H_o : (CD)_{11} = \cdots = (CD)_{cd} = 0$$

$$H_o : (CE)_{11} = \cdots = (CE)_{ce} = 0$$

$$H_o : (DE)_{11} = \cdots = (DE)_{de} = 0$$

$$H_o : (BCD)_{111} = \cdots = (BCD)_{bcd} = 0$$

$$H_o : (BCE)_{111} = \cdots = (BCE)_{bce} = 0$$

$$H_o : (BDE)_{111} = \cdots = (BDE)_{bde} = 0$$

$$H_o : (CDE)_{111} = \cdots = (CDE)_{cde} = 0$$

$$H_o : (BCDE)_{1111} = \cdots = (BCDE)_{bcde} = 0$$

Según Johnson y Leone (1964), la tabla de análisis de varianza es como sigue

## TABLA 1. ANALISIS DE VARIANZA

| Fuente | Suma de Cuadrados | Cuadrados Medios Esperados |
|---|---|---|
| B | $\sum_t \frac{y^2_{t....}}{cden} - \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{cden}{b-1} \sum_t B_t^2$ |
| C | $\sum_i \frac{y^2_{.i...}}{bden} - \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{bden}{c-1} \sum_i C_i^2$ |
| D | $\sum_j \frac{y^2_{..j..}}{bcen} - \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{bcen}{d-1} \sum_j D_j^2$ |
| E | $\sum_k \frac{y^2_{...k.}}{bcdn} - \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{bcdn}{e-1} \sum_k E_k^2$ |
| BC | $\sum_{ti} \frac{y^2_{ti...}}{den} - \sum_t \frac{y^2_{t....}}{cden} - \sum_i \frac{y^2_{.i...}}{bden} + \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{den}{(b-1)(c-1)} \sum_{ti} (BC)_{ti}^2$ |
| BD | $\sum_{tj} \frac{y^2_{t.j..}}{cen} - \sum_t \frac{y^2_{t....}}{cden} - \sum_j \frac{y^2_{..j..}}{bcen} + \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{cen}{(b-1)(d-1)} \sum_{tj} (BD)_{tj}^2$ |
| BE | $\sum_{tk} \frac{y^2_{t..k.}}{cdn} - \sum_t \frac{y^2_{t....}}{cden} - \sum_k \frac{y^2_{...k.}}{bcdn} + \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{cdn}{(b-1)(e-1)} \sum_{tk} (BE)_{tk}^2$ |
| CD | $\sum_{ij} \frac{y^2_{.ij..}}{ben} - \sum_i \frac{y^2_{.i...}}{bden} - \sum_j \frac{y^2_{..j..}}{bcen} + \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{ben}{(c-1)(d-1)} \sum_{ij} (CD)_{ij}^2$ |
| CE | $\sum_{ik} \frac{y^2_{.i.k.}}{ben} - \sum_i \frac{y^2_{.i...}}{bden} - \sum_k \frac{y^2_{...k.}}{bcdn} + \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{bdn}{(c-1)(e-1)} \sum_{ik} (CE)_{ik}^2$ |
| DE | $\sum_{jk} \frac{y^2_{..jk.}}{bcn} - \sum_j \frac{y^2_{..j..}}{bcen} - \sum_k \frac{y^2_{...k.}}{bcdn} + \frac{y^2_{.....}}{bcden}$ | $\sigma^2 + \frac{bcn}{(d-1)(e-1)} \sum_{jk} (DE)_{jk}^2$ |
| BCD | $\sum_{tij} \frac{y^2_{tij..}}{-} \sum_{ti} \frac{y^2_{ti...}}{den} - \sum_{tj} \frac{y^2_{t.j..}}{cen} - \sum_{ij} \frac{y^2_{.ij..}}{ben} +$ | $\sigma^2 + \frac{en}{(b-1)(c-1)(d-1)} \sum_{tij} (BCD)_{tij}^2$ |
| | $\sum_t \frac{y^2_{t....}}{cden} + \sum_i \frac{y^2_{.i...}}{bden} + \sum_j \frac{y^2_{..j..}}{bcen} - \frac{y^2_{.....}}{bcden}$ | |
| BCE | $\sum_{tik} \frac{y^2_{ti.k.}}{dn} - \sum_{ti} \frac{y^2_{ti...}}{den} - \sum_{tk} \frac{y^2_{t..k.}}{cdn} - \sum_{ik} \frac{y^2_{.i.k.}}{bdn} +$ | $\sigma^2 + \frac{dn}{(b-1)(c-1)(e-1)} \sum_{tik} (BCE)_{tik}^2$ |
| | $\sum_t \frac{y^2_{t....}}{cden} + \sum_i \frac{y^2_{.i...}}{bden} + \sum_k \frac{y^2_{...k.}}{bcen} - \frac{y^2_{.....}}{bcden}$ | |

(... continuación de la Tabla 1)

| Fuente | Suma de Cuadrados | Cuadrados Medios Esperados |
|---|---|---|
| BDE | $\sum_{tjk} \frac{y^2_{t.jk.}}{cn} - \sum_{tj} \frac{y^2_{t.j..}}{cen} - \sum_{tk} \frac{y^2_{t..k.}}{cdn} - \sum_{jk} \frac{y^2_{..jk.}}{bcn} +$ | $\sigma^2 +$ |
| | $\sum_{t} \frac{y^2_{t....}}{cden} + \sum_{j} \frac{y^2_{..j..}}{bcen} + \sum_{k} \frac{y^2_{...k.}}{bcen} - \frac{y^2_{.....}}{bcden}$ | $\frac{cn}{(b-1)(d-1)(e-1)} \sum_{tjk} (BDE)^2_{tjk}$ |
| CDE | $\sum_{ijk} \frac{y^2_{.ijk.}}{bn} - \sum_{ij} \frac{y^2_{.ij..}}{ben} - \sum_{ik} \frac{y^2_{.i.k.}}{bdn} - \sum_{jk} \frac{y^2_{..jk.}}{bcn} +$ | $\sigma^2 +$ |
| | $\sum_{i} \frac{y^2_{.i...}}{bden} + \sum_{j} \frac{y^2_{..j..}}{bcen} + \sum_{k} \frac{y^2_{...k.}}{bcen} - \frac{y^2_{.....}}{bcden}$ | $\frac{bn}{(c-1)(d-1)(e-1)} \sum_{ijk} (CDE)^2_{ijk}$ |
| BCDE | $\sum_{tijk} \frac{y^2_{tijk.}}{n} - \sum_{tij} \frac{y^2_{tij..}}{en} - \cdots + \sum_{ti} \frac{y^2_{ti...}}{den} + \cdots$ | $\sigma^2 +$ |
| | $- \sum_{t} \frac{y^2_{t....}}{cden} - \cdots + \frac{y^2_{.....}}{bcden}$ | $\frac{n}{(b-1)(c-1)(d-1)(e-1)} \sum_{tijk} (BCDE)^2_{tijk}$ |
| Residual | Diferencia | $\sigma^2$ |
| Total | $\sum_{tijkl} y^2_{tijkl} - \frac{y^2_{.....}}{bcden}$ | |

donde los grados de libertad de cada factor aparecen en el denominador del coeficiente que multiplica al segundo sumando del esperado de los cuadrados medios.

También es posible estimar cada uno de los términos del modelo (1.1). Como lo muestra Graybill (1961), los estimadores vienen dados por

$$\hat{A} = \bar{y}_{.....};$$

$$\hat{B}_t = \bar{y}_{t....} - \bar{y}_{.....};$$

$$\hat{C}_i = \bar{y}_{.i...} - \bar{y}_{.....};$$

$$\hat{D}_j = \bar{y}_{..j..} - \bar{y}_{.....};$$

$$\hat{E}_k = \bar{y}_{...k.} - \bar{y}_{.....};$$

$$(\hat{BC})_{ti} = \bar{y}_{ti...} - \bar{y}_{t....} - \bar{y}_{.i...} + \bar{y}_{.....}$$

$$(\hat{BD})_{tj} = \bar{y}_{t.j..} - \bar{y}_{t....} - \bar{y}_{..j..} + \bar{y}_{.....}$$

$$(\hat{BE})_{tk} = \bar{y}_{t..k.} - \bar{y}_{t....} - \bar{y}_{...k.} + \bar{y}_{.....}$$

$$(\hat{CD})_{ij} = \bar{y}_{.ij..} - \bar{y}_{.i...} - \bar{y}_{..j..} + \bar{y}_{.....}$$

$$(\hat{CE})_{ik} = \bar{y}_{.i.k.} - \bar{y}_{.i...} - \bar{y}_{...k.} + \bar{y}_{.....}$$

$$(\hat{DE})_{jk} = \bar{y}_{..jk.} - \bar{y}_{..j..} - \bar{y}_{...k.} + \bar{y}_{.....}$$

$$(\hat{BCD})_{tij} = \bar{y}_{tij..} - \bar{y}_{ti...} - \bar{y}_{t.j..} - \bar{y}_{.ij..} + \bar{y}_{t....} + \bar{y}_{.i...} + \bar{y}_{..j..} - \bar{y}_{.....}$$

$$(\hat{BCE})_{tik} = \bar{y}_{ti.k.} - \bar{y}_{ti...} - \bar{y}_{t..k.} - \bar{y}_{.i.k.} + \bar{y}_{t....} + \bar{y}_{.i...} + \bar{y}_{...k.} - \bar{y}_{.....}$$

$$(\hat{BDE})_{tjk} = \bar{y}_{t.jk.} - \bar{y}_{t.j..} - \bar{y}_{t..k.} - \bar{y}_{..jk.} + \bar{y}_{t....} + \bar{y}_{..j..} + \bar{y}_{...k.} - \bar{y}_{.....}$$

$$(\hat{CDE})_{ijk} = \bar{y}_{.ijk.} - \bar{y}_{.ij..} - \bar{y}_{.i.k.} - \bar{y}_{..jk.} + \bar{y}_{.i...} + \bar{y}_{..j..} + \bar{y}_{...k.} - \bar{y}_{.....}$$

$$(\hat{BCDE})_{tijk} = \bar{y}_{tijk.} - \bar{y}_{tij..} - \bar{y}_{ti.k.} - \bar{y}_{t.jk.} - \bar{y}_{.ijk.} + \bar{y}_{ti...} + \bar{y}_{t.j..} + \bar{y}_{t..k.} + \bar{y}_{.ij..}$$
$$+ \bar{y}_{.i.k.} + \bar{y}_{..jk.} - \bar{y}_{t....} - \bar{y}_{.i...} - \bar{y}_{..j..} - \bar{y}_{...k.} + \bar{y}_{.....}$$

así como los cuadrados medios esperados para $\sigma^2$. Para poder realizar la prueba de las hipótesis (1.4) es fundamental el supuesto de normalidad (1.3 d).

## 2. Modelo con Efectos Fijos Cruzados Balanceados sin la Hipótesis de Normalidad

Nuevamente consideraremos un modelo de 4 factores. Debido a que la normalidad de $\epsilon_{tijkl}$ implica la normalidad de $y_{tijkl}$, (1.1) y (1.3) son equivalentes a

$$\mu_{tijkl} = A + B_t + C_i + D_j + E_k + \tag{2.1}$$

$$+ (BC)_{ti} + (BD)_{tj} + (BE)_{tk} + (CD)_{ij} + (CE)_{ik} + (DE)_{jk} +$$

$$+ (BCD)_{tij} + (BCE)_{tik} + (BDE)_{tjk} + (CDE)_{ijk} + (BCDE)_{tijk}$$

$$t = 1, \cdots, b; \quad i = 1, \cdots, c; \quad j = 1, \cdots, d; \quad k = 1, \cdots, e; \quad l = 1, \cdots, n$$

y

$$E(y_{tijkl}) = \mu_{tijk} \tag{2.2a}$$

$$Var(y_{tijkl}) = \sigma^2 \tag{2.2}$$

$$Cov(y) = \sigma^2 \; I \tag{2.2c}$$

$$y_{tijkl} \sim N(\mu_{tijkl}, \sigma^2) \tag{2.2d}$$

El análisis de varianza utiliza de forma esencial la suma de cuadrados y descompone esta en sumas de cuadrados independientes, de manera que en la medida en que la suma de cuadrados del error sea menor, mejor ajustará el modelo (1.1). Por ello, suele minimizarse la suma de cuadrados en muchas de las aplicaciones estadísticas en donde no se hacen supuestos a cerca de las distribuciones del error. Pero, minimizar cuadrados (en caso de suponer una distribución para el error) es equivalente a suponer una distribución normal y obtener estimadores de máxima verosimilitud.

El supuesto de normalidad para el error suele ser muy natural en muchas aplicaciones pero en otras no. Así surge la necesidad de un análisis de varianza para distribuciones no normales (en caso de tener no normalidad debemos cambiar 2.2 c) por 2.2d) las $y_{tijkl}$ son independientes entre sí).

En lo que sigue se mostrará que el análisis de desvío es una herramienta análoga al análisis de varianza pero más flexible en dos aspectos:

(1) Permite que la variable dependiente tenga una gama mucho más amplia de distribuciones (entre las cuales está incluida la normal y en ese caso el análisis del desvío se reduce al análisis de varianza clásico.

(2) Provee de mayor cantidad de pruebas de hipótesis con una mayor simplicidad.

Supongamos que $Y$ es una variable aleatoria no degenerada con función de distribución de probabilidades dada por

$$p(y;\theta,\lambda) = a(y;\lambda)\ e^{\lambda(y\theta-\kappa(\theta))} \tag{2.3}$$

donde $\theta$ y $\lambda$ son los parámetros de la distribución con $(\theta,\lambda) \in \Theta \times \Lambda$, con $\Theta = \{\theta \in \mathbb{R};\ \int_{supp\ a} a(y;\lambda)\ e^{\lambda y\theta}\ d\theta < \infty\}$ y con $a, \kappa$ funciones suficientemente diferenciables. Jorgensen (1992) demuestra que $\Theta$ es un intervalo y que $\Lambda$ es un semigrupo aditivo, el cual coincide con $\Lambda = \mathbb{R}^+$ o $\Lambda = \mathbb{Z}^+$ si se supone además que $\Lambda \subset \mathbb{R}$ y que $1 \in \Lambda$.

Se demuestra fácilmente que las distribuciones normal, binomial, Poisson, binomial negativa, gama (chi cuadrado, exponencial), secante hipérbolica Generalizada; la Poisson compuesta con gama, las generadas por extremo estables (ver Feller, 1971) (en donde generada significa que $a$ es la extremo estable y $\kappa$ es tal que 2.3 sea una distribución de probabilidades) y aquellas generadas por cualquier distribución no degenerada pueden escribirse de la forma (2.3).

Jorgensen (1992) también prueba que si *int* $\Theta$ es diferente de vacío y $Y$ no tiene una distribución degenerada entonces $E(Y) = \lambda\kappa'(\theta)$ y $Var(Y) = \lambda\kappa''(\theta)$. Como la variable aleatoria es no degenerada $Var(Y) > 0$ y por lo tanto $\tau = \kappa'$ es monótona creciente y tiene inversa. Si hacemos $\mu = \tau(\theta)$ entonces $\theta = \tau^{-1}(\mu)$ y en ese caso $E(Y) = \lambda\mu$ y $Var(Y) = \lambda V(\mu)$ donde $V(\mu) = \kappa''(\tau^{-1}(\mu)) > 0$. La función $V$ es importante porque indica cual es la relación funcional que hay entre la varianza y el valor esperado de la variable supuesto $\lambda$ conocido y se denomina función de varianza. Un resultado muy importante dice que las funciones de varianza caracterizan una distribución dentro de la familia de dispersión exponencial.

Si hacemos $\sigma^2 = 1/\lambda$ y dado que $\theta = \tau^{-1}(\mu)$ entonces (2.3) puede escribirse como, Jorgensen (1992),

$$p(y;\mu,\sigma^2) = a(y;\sigma^{-2})\ e^{\sigma^{-2}(y\tau^{-1}(\mu)-\kappa(\tau^{-1}(\mu)))} \tag{2.4}$$

En esta expresión se han cambiado los parámetros $(\theta,\lambda) \in \Theta \times \Lambda$ por $(\mu,\sigma^2) \in \Omega \times \Sigma$ donde $\Omega = \tau(Int\Theta)$ y $\Sigma = \Lambda^{-1}$. A la familia (2.4) se le denota por $ED(\mu,\sigma^2)$.

Supongamos que se toma una muestra aleatoria independiente de tamaño $n$ de de variables aleatorias $Y_i$ que toman valores $y_i$ de acuerdo a

$$a(y_i; w_i\sigma^{-2})\ e^{w_i\sigma^{-2}t(y_i,\mu_i)}$$

donde $t(y_i,\mu_i) = y_i\tau^{-1}(\mu_i) - \kappa(\tau^{-1}(\mu_i))$ y con los $w_i$, $i = 1,\cdots,n$ conocidos. Entonces, el logaritmo de la función de verosimilitud viene dado por

$$ln\,L(\mu,\sigma^2;y) = \sum_{i=1}^{n} ln\,a(y_i;w_i\sigma^{-2}) + \sigma^{-2}\sum_{i=1}^{n} w_i t(y_i,\mu_i). \tag{2.5}$$

Se define la función de desvío del vector de parámetros $\mu = (\mu_1, \cdots, \mu_n)$ dado el vector de observaciones $y = (y_1, \cdots, y_n)$ por

$$D(\mu; y) = 2\sum_{i=1}^{n} d_i^2 = 2\sum_{i=1}^{n} w_i \left( sup_{\mu_i^* \in \Omega} t(y_i, \mu_i^*) - t(y_i, \mu_i) \right) \tag{2.6}$$

Esta definición es en realidad $-2\, ln\, R$, donde $R$ es la razón de verosimilitud entre el modelo obtenido a partir de los $\mu_i$, $i = 1, \cdots, n$ y el modelo obtenido a partir de los $\mu_i^*$, $i = 1, \cdots, n$, donde los $\mu_i^*$ son los que mejor explican la muestra aleatoria $y_i$.

Se observa fácilmente que $D(\mu; y) \geq 0$ pues cada $d_i^2 \geq 0$ ya que $\mu_i^*$ es el que hace "máximo" $t(y_i, \mu_i)$ conocido $y_i$ mientras que $t(y_i, \mu_i)$ es el valor de $t$ conocidos $\mu_i$ y $y_i$.

Por otro lado (2.6) muestra la naturaleza aditiva del desvío.

De (2.5) se tiene que

$$ln\, L(\mu, \sigma^2; y) = c(\sigma^2; y) - \frac{D(\mu; y)}{2\sigma^2} \tag{2.7}$$

con

$$c(\sigma^2; y) = \sum_{i=1}^{n} \left( ln\, a(y_i; w_i \sigma^{-2}) + w_i \sigma^{-2} sup_{\mu_i^* \in \Omega} t(y_i, \mu_i^*) \right) \tag{2.8}$$

Luego, de (2.7), maximizar $ln\, L(\mu, \sigma^2; y)$ es equivalente a minimizar $D(\mu; y)$ ya que $c(\sigma^2; y)$ es independiente de $\mu$ y al derivar (2.5) respecto de $\mu$ e igualar a cero (para encontrar los estimadores de máxima verosimilitud) se obtiene una expresión independiente de $\sigma^2$.

El cálculo del término $sup_{\mu_i^* \in \Omega} t(y_i, \mu_i^*)$ se puede realizar de la siguiente manera. Considere

$$l(\theta) = t(y, \mu) = t(y, \tau(\theta)) = y\theta - \kappa(\theta) \tag{2.9}$$

entonces

$$l'(\theta) = y - \kappa'(\theta) = y - \tau(\theta) \tag{2.10}$$

$$l''(\theta) = -\kappa''(\theta) = -V(\mu) < 0 \tag{2.11}$$

Por lo tanto, de la concavidad estricta de $l$, ella posee un único máximo el cual ocurre en $\tilde{\theta}$. Si $\tilde{\theta} \in Int\Theta$ entonces igualando a cero (2.10) se tiene

$$\tilde{\theta} = \tau^{-1}(y) \tag{2.12}$$

Como para una variable aleatoria continua, la frontera de $\Theta$ tiene probabilidad cero, entonces (2.12) existe con probabilidad 1. Lo mismo para para las distribuciones mixtas que son miembros de la familia de dispersión exponencial. En cambio, para el caso discreto, la frontera tiene una probabilidad positiva y puede ocurrir que (2.12) no tenga solución. Sin embargo, para los casos discretos que más se presentan en las aplicaciones como Poisson, binomial, binomial negativa ese no es el caso y por lo tanto (2.12) existe.

Entonces, es posible expresar el desvío de la siguiente manera

$$D(\mu; y) = 2\sum_{i=1}^{n} w_i \Big( (y_i\tilde{\theta}_i - \kappa(\tilde{\theta}_i) - (y_i\theta_i - \kappa(\theta_i)) \Big) \tag{2.13}$$

Si $Y \sim N(\mu_Y, \sigma^2)$ entonces

$$(2\pi\sigma^2)^{-1/2}\, e^{-\frac{1}{2\sigma^2}(Y-\mu_Y)^2} = (2\pi\sigma^2)^{-1/2}\, e^{-\frac{Y}{2\sigma^2}}\, e^{Y\mu_Y - \frac{1}{2}\mu_Y^2}$$

y en este caso $\theta = \mu_Y$, $\kappa(\theta) = \frac{1}{2}\theta^2$ y $a(y;\sigma^2) = (2\pi\sigma^2)^{-1/2}\, e^{-\frac{Y}{2\sigma^2}}$. Luego, $\kappa' = \tau = \tau^{-1} = Id$; $\mu = \mu_Y$; $V(\mu) = 1$ y $t(y,\mu) = y\mu - \frac{1}{2}\mu^2$.

Si se toma una muestra aleatoria de tamaño $n$ entonces $Y \sim N(\mu, \sigma^2 W)$ donde $Y = (Y_1, \cdots, Y_n)$; $\mu = (\mu_1, \cdots, \mu_n)$ y $W = diag\{w_i^{-1}\}$, de (2.13) se tiene que

$$D(\mu; y) = 2\sum_{i=1}^{n} w_i((y_i^2 - \frac{1}{2}y_i^2) - (y_i\mu_i - \frac{1}{2}\mu_i^2))$$

$$= 2\sum_{i=1}^{n} w_i(\frac{1}{2}y_i^2 - y_i\mu_i + \frac{1}{2}\mu_i^2)$$

$$= \sum_{i=1}^{n} w_i(y_i - \mu_i)^2 \qquad (2.14)$$

que es la suma de cuadrados pesados con pesos $w_i$ y si los $w_i = 1$, $i = 1, \cdots, n$ entonces se obtiene la suma de cuadrados considerada en el análisis de varianza con distribución normal para $Y$.

Como se mencionó anteriormente, maximizar la función de verosimilitud es equivalente a minimizar la función desvío. Una ventaja de utilizar esta última estadística reside en la posibilidad de cambiar de distribución de la variable aleatoria dependiente (dentro de la amplia familia de modelos de dispersión exponencial) y comparar los desvíos calculados para cada uno de esos modelos, obteniendo como mejor ajuste aquel modelo para el cual se obtenga el menor desvío.

En la página que sigue se muestra una tabla, tomada de Jorgensen (1992), página 65 con los desvíos de las distribuciones miembros de la familia de dispersión exponencial más relevantes.

Si denotamos por $D_1$ el desvío de un modelo con $k_1$ grados de libertad (donde $k_1 = n - p_1$) con el que se está trabajando y por $D_2$ el desvío de un submodelo del anterior con $k_2$ grados de libertad (donde $k_2 = n - p_2$ y $k_2 \leq k_1$ lo que implica $p_1 \geq p_2$), es decir, que tiene sólo un subconjunto de los parámetros que aparecen en el primer modelo y por $w = min\{w_1, \cdots, w_n\}$, y si entendemos por $w \to \infty$ que todas las $w_i$ son suficientemente grandes entonces puede demostrarse que

$$\frac{D_1}{\sigma^2} \sim \chi^2_{(p_1)} \quad \text{si} \quad w \to \infty \tag{2.15}$$

$$\frac{D_2 - D_1}{\sigma^2} \sim \chi^2_{(k_1 - k_2)} \quad \text{si} \quad w \to \infty \quad \text{o si} \quad n \to \infty \tag{2.16}$$

TABLA 2

DESVIOS EN MODELOS

| Modelo | Desvío |
|---|---|
| Normal | $\sum_{i=1}^{n} w_i(y_i - \mu_i)^2$ |
| Poison | $2\sum_{i=1}^{n} w_i\left(y_i \, ln \, \frac{y_i}{\mu_i} - (y_i - \mu_i)\right)$ |
| Binomial | $2\sum_{i=1}^{n} w_i\left(y_i \, ln \, \frac{y_i}{\mu_i} - (m - y_i) \, ln \left(\frac{m - y_i}{m - \mu_i}\right)\right)$ |
| Binomial Negativa | $2\sum_{i=1}^{n} w_i\left(y_i \, ln \, \frac{y_i}{\mu_i} + (y_i + r) \, ln \left(\frac{1 + \mu_i}{1 + y_i}\right)\right)$ |
| Gamma | $2\sum_{i=1}^{n} w_i\left(\frac{y_i}{\mu_i} - 1 + ln \, \frac{\mu_i}{y_i}\right)$ |
| Secante Hiperbólica Generalizada | $2\sum_{i=1}^{n} w_i\left(y_i \, arctan \left(\frac{y_i - \mu_i}{1 + y_i\mu_i}\right) + \frac{1}{2} \, ln \left(\frac{1 + \mu_i^2}{1 + y_i^2}\right)\right)$ |
| Inversa Gaussiana | $\sum_{i=1}^{n} \frac{w_i}{y_i\mu_i^2} (y_i - \mu_i)^2$ |
| Modelo con Varianza $\mu^p$ $p \neq 0, 1, 2, 3$; $\alpha = \frac{p-2}{p-1}$ | $2(\alpha - 1)\sum_{i=1}^{n} w_i\left(\frac{\alpha-1}{\alpha} \, y_i^{\frac{\alpha}{\alpha-1}} + \mu_i^{\frac{1}{\alpha-1}} \left(\frac{\mu_i}{\alpha} - y_i\right)\right)$ |
| Modelo con Varianza Exponencial | $2\sum_{i=1}^{n} w_i e^{\mu_i}\left(e^{y_i - \mu_i} - 1 - (y_i - \mu_i)\right)$ |

Debido a que no se conocen los parámetros $\mu$ del modelo, el desvío tiene que ser estimado. Para ello, si relacionamos cada uno de los parámetros $\mu$ del modelo ($n$ en total) con $p$ parámetros $\beta$ con $p < n$ entonces $\mu$ dependerá de $\eta$ y se tendrá $\mu = \mu(\beta)$ y por lo tanto $D(\mu; y) = D(\mu(\beta); y)$; es decir, el desvío dependerá exclusivamente de los nuevos parámetros $\beta$.

Luego, una vez estimado $\beta$ se tendrá una estimación del desvío y a partir de ahí, considerando (2.5) será posible obtener una estimación para $\sigma^2$ mediante procedimientos iterativos.

Por otro lado, es importante observar que las distribuciones discretas más importantes sólo poseen un parámetro (por ejemplo, la Poisson) y en ese caso tiene que hacerse $\lambda = 1$ en (2.3) o el parámetro $\lambda$ tiene un valor conocido de antemano (por ejemplo, en la binomial se tendrá que $\lambda = n$, donde $n$ es el número de ensayos preestablecido).

En ambos casos, carece de sentido realizar una estimación de $\sigma^2$.

## 3. Modelo con Efectos Fijos Cruzados Balanceados sin la Hipótesis de Normalidad en el Contexto de los Modelos Lineales Generalizados

Consideremos nuevamente (2.1)

$$\begin{aligned}\mu_{tijkl} = {} & A + B_t + C_i + D_j + E_k + \\ & + (BC)_{ti} + (BD)_{tj} + (BE)_{tk} + (CD)_{ij} + (CE)_{ik} + (DE)_{jk} + \\ & + (BCD)_{tij} + (BCE)_{tik} + (BDE)_{tjk} + (CDE)_{ijk} + (BCDE)_{tijk}\end{aligned} \tag{3.1}$$

$$t = 1, \cdots, b; \quad i = 1, \cdots, c; \quad j = 1, \cdots, d; \quad k = 1, \cdots, e; \quad l = 1, \cdots, n$$

pero sin los supuestos (2.2) sino solamente con

$$\text{Las } y_{tijkl} \text{ son independientes entre sí} \tag{3.2a}$$

$$y_{tijkl} \sim ED(\mu_{tijkl}, \sigma^2/w_{tijkl}) \text{ con los } w_{tijkl} \text{ conocidos.} \tag{3.2b}$$

Es muy interesante observar que si hacemos $l = 1$ en (3.1) y se supone una distribución Poisson para $Y$ en 3.2 b) tenemos el modelo logístico para tablas de contingencia.

Por un modelo lineal generalizado se entiende un modelo que consta de dos componentes. Una aleatoria que especifica la distribución de la variable dependiente siendo

$$Z_s \sim ED(\mu_s, \sigma^2/w_s), \qquad s = 1, \cdots, n \tag{3.3}$$

con los $w_s$ conocidos y donde además se supone que las $Z_s$ son independientes entre sí y una componente sistemática que relaciona el valor esperado de la variable dependiente con los parámetros $\beta$ del modelo a través de una matriz de diseño $X$ (constante y conocida) y de una función $g$ de enlace inversible, suficientemente diferenciable de la siguiente manera

$$g(\mu_s) = \eta_s = x_{s0}\beta_0 + x_{s1}\beta_1 + \cdots + x_{sp}\beta_p \tag{3.4}$$

o equivalentemente $\mu_s = g^{-1}(\eta_s)$.

El algoritmo de estimación de los parámetros $\beta = (\beta_0, \cdots, \beta_p)$ sigue el proceso iterativo denominado método score para parámetros de Fisher, el cual Dumett (1995) lo expresa de la siguiente manera

$$\beta^{(m+1)} = (X^t H^{(m)} W X)^{-1} X^t H^{(m)} W y^{*(m)} \tag{3.5}$$

donde

$$
\begin{aligned}
y^* &= X\beta + G(y-\mu)\\
G &= diag\{g'(\mu_s)\}\\
H &= diag\{h_s\} \qquad \text{con} \quad h_s = \frac{1}{V(\mu_s)}\left(\frac{d\mu_s}{d\eta_s}\right)^2\\
W &= diag\{w_s\}
\end{aligned}
$$

en donde $(m)$ indica la estimación en la $m-$ésima iteración. A $y^*$ se le denomina variable dependiente modificada. Debido a (3.4), $\mu_s$ depende de $\beta$ entonces $H$, $G$, $y^*$ dependen de $\beta$. El proceso iterativo se inicia con $\hat{\mu}_s^o = y_s$, es decir se estima el valor promedio por el dato observado.

Es fácil ver que (3.1) y (3.2) pueden ser vistos como un modelo lineal generalizado si (3.1) se considera la componente sistemática con $g = Id$ (lo que implica que $G = I_n$ y $H = diag\{V^{-1}(\mu_s)\}$) y donde $\beta_0 = A$; $\beta_1 = B_1; \cdots$ ; $\beta_t = B_t$; $\beta_{t+1} = C_1; \cdots$ ; $\beta_{t+i} = C_i$; $\beta_{t+i+1} = D_1; \cdots$ y asísucesivamente donde $p = t + i + j + k + ti + tj + tk + ij + ik + jk + tij + tik + tjk + ijk + tijk$ y $n = tijkl$ y la componente aleatoria viene dada por (3.2).

En caso de especificar la distribución normal para $Y$ se tendrá $H = I_n$,

$y^* = y$ pues $\mu = X\beta$ y el esquema iterativo se reducirá a un algoritmo no iterativo para $\hat{\beta}$ dado por

$$\hat{\beta} = (X^tWX)^{-1}X^tWy \tag{3.6}$$

los cuales son conocidos como los estimadores de máxima verosimilitud ponderados por $W$ para $\beta$.

Como estimador de $\sigma^2$ se toma más frecuentemente

$$\hat{\sigma}^2 = \frac{D(\mu;y)}{n-p} \tag{3.7}$$

En Jorgensen (1992) también existen resultados distribucionales asociados a los estimadores de $\beta$ y $\sigma^2$ entre los cuales se encuentran

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^t \hat{H}(\beta) W X)^{-1}), \quad \text{si} \quad w \to \infty \quad \text{o si} \quad n \to \infty \tag{3.8}$$

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \left(X^t \hat{H}(\beta) W X\right)^{-1}_{jj}}} \sim t_p, \quad \text{si} \quad w \to \infty \quad \text{o si} \quad n \to \infty \tag{3.9}$$

$$\frac{(n - k_1)\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-k_1)}, \quad \text{si} \quad w \to \infty \tag{3.10}$$

$$\frac{D_2 - D_1}{\hat{\sigma}^2\,(k_1 - k_2)} \sim F(k_1 - k_2, n - k_1) \quad \text{si} \quad w \to \infty \quad \text{o si} \quad n \to \infty \tag{3.11}$$

en donde $D_1$, $D_2$, $k_1$, $k_2$ tienen los mismos significados de antes.

Se puede observar el parecido notable con las fórmulas de distribución para los modelos de regresión normal multivariado.

Es importante observar que no pueden utilizarse siempre estos resultados distribucionales pues a parte de tener que verificar las condiciones de ser resultados límite cuando $w \to \infty$ o cuando $n \to \infty$, no se conoce necesariamente el valor de $\sigma^2$. Pero, $\sigma^2$ si es conocido en la mayoría de las distribuciones discretas que son miembros de la familia de dispersión exponencial y en esos casos es posible realizar un estudio más completo.

Como se mencionó anteriormente, el desvío es de naturaleza aditiva. Esto nos permite realizar un análisis del desvío análogo al análisis de varianza. Teniendo en cuenta los resultados distribucionales (2.15)-(2.16) y (3.8)-(3.11) la tabla de análisis de desvío se formula de la siguiente manera (tomada de Jorgensen, (1992), pág 98).

TABLA 3. ANALISIS DEL DESVIO

| Modelo | Desvío | g.l. | $\Delta D_i$ | $\Delta$ g.l. | $\hat{\sigma}^2$ | F |
|---|---|---|---|---|---|---|
| $H_1$ | $D_1$ | $f_1$ | | | $\hat{\sigma}^2_{(1)}$ | |
| $H_2$ | $D_2$ | $f_2$ | $D_2 - D_1$ | $f_2 - f_1$ | $\hat{\sigma}^2_{(2)}$ | $F_2$ |
| $H_3$ | $D_3$ | $f_3$ | $D_3 - D_2$ | $f_3 - f_2$ | $\hat{\sigma}^2_{(3)}$ | $F_3$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

donde $H_1 \supseteq H_2 \supseteq \cdots$ es una sucesión de hipótesis anidadas (modelos), $H_s$ con desvío $D_s$ y $f_s$ grados de libertad. En esta tabla, $D$ significa desvío; $f$ grados de libertad; $\Delta D_i = D_i - D_{i-1}$, es la diferencia de dos desvíos del modelo y del modelo anterior; $\Delta g.l. = f_i - f_{i-1}$, es la diferencia de grados de libertad del modelo y del modelo anterior; $\hat{\sigma}^2$ es el estimador de $\sigma^2$ dado por $D(\mu(\beta); y)/f$ y $F$ es el valor de la estadística

$$F_i = \frac{(D_i - D_{i-1})/(f_i - f_{i-1})}{\hat{\sigma}^2_{(i-1)}}$$

la cual debe compararse contra el valor crítico de una distribución $F$ con $f_i - f_{i-1}$ y $f_{i-1}$ grados de libertad para docimar $H_i$ bajo $H_{i-1}$ de la manera siguiente : si se rechaza la prueba se acepta $H_i$. Se empieza probando $H_2$ bajo $H_1$.

## 4. Ejemplos

En esta sección se presentan dos ejemplos. El primero de ellos fué tomado de un texto de análisis de varianza clásico y por lo tanto es posible desarrollar un análisis tradicional del mismo. Dicho ejemplo también se analiza con los mismos supuestos pero dentro del contexto de los modelos lineales generalizados de manera que se puedan comparar ambas metodologías de trabajo. Una vez dentro del marco de los modelos lineales generalizados se ajustarán los datos suponiendo otras distribuciones para la variable dependiente y se encontrarán mejores resultados que con la distribución normal.

El segundo ejemplo fué creado por el autor del presente trabajo, utilizando simulación de distribuciones normales, con el objeto de estudiar que ocurre en esos casos con los modelos de distribuciones alternativas.

Todos los ejemplos han sido desarrollados con el paquete GLIM (Generalized Linear Interactive Modeling), el cual fué desarrollado por la Royal Statistical Society específicamente para el análisis de modelos lineales generalizados. Este software es el único en su género y en estos momentos SAS posee un procedimiento llamado GENMOD para trabajar con modelos lineales generalizados, el cual se encuentra en estado de experimentación.

**Ejemplo 1.** Este problema ha sido tomado de Scheffé (1959) pág 145. Este es un experimento con 4 factores. Las observaciones son el contenido de una mezcla (en gramos) de un cierto producto alimenticio en un estado de experimentación. Los niveles del factor $A$ son tres tipos de sal; los de $B$ corresponden a las cantidades de esas sales (las mismas cantidades molares para cada sal); las de $C$ son cantidades de cierto ácido y las de $D$ son dos diferentes aditivos (dice el texto, que esos datos fueron proporcionados por Mr. Otto Dykstra Jr. y reproducido con el permiso del Research Center of General Foods Corporation). Los datos y las tablas de análisis de varianza y de desvío para este modelo se presentan en las páginas siguientes.

En la tabla de análisis de desvío se comenzó por el modelo completo y se fue quitando en cada paso un término al modelo (sin que necesariamente sea un procedimiento backward) para mostrar la naturaleza secuencial de este procedimiento. El objetivo primordial de haber realizado esto es mostrar que ambos análisis coinciden para el caso de distribución normal. Una vez entendido como funciona la tabla de análisis de desvío, procedemos a ajustar estos modelos con otras distribuciones.

En el apéndice se muestra la salida del programa GLIM para estos mismos datos suponiendo distribuciones normal, gama, Poisson e inversa Gaussiana y para cada una de esos modelos se plantean 4 modelos : primero el que contiene las interacciones de orden tres, luego el que incluye sólo interacciones de orden dos, después el que no tiene interacciones y finalmente el modelo de únicamente la constante (salvo en el caso normal en que no se incluye este último debido a que el modelo sin interacciones es insuficiente para ajustar los datos). En el apéndice puede encontrarse el programa GLIM utilizado.

TABLA 4. DATOS DEL EJEMPLO 1

| | | NIVEL DE C | | | |
|---|---|---|---|---|---|
| NIVEL DE A | NIVEL DE B | 1 | | 2 | |
| | | NIVEL DE D | | NIVEL DE D | |
| | | 1 | 2 | 1 | 2 |
| 1 | 1 | 8 | 5 | 8 | 4 |
| | 2 | 17 | 11 | 13 | 10 |
| | 3 | 22 | 16 | 20 | 15 |
| 2 | 1 | 7 | 3 | 10 | 5 |
| | 2 | 26 | 17 | 24 | 19 |
| | 3 | 34 | 32 | 34 | 29 |
| 3 | 1 | 10 | 5 | 9 | 4 |
| | 2 | 24 | 14 | 24 | 16 |
| | 3 | 39 | 33 | 36 | 34 |

Los resultados se pueden observar en la tabla que se encuentra en la página 74. En ella, $M_o$ es el modelo que incluye todos los términos del modelo incluídas las interacciones entre tres factores; $M_1$ incluye todas las interacciones; $M_3$ es el modelo sin interacciones y $M_4$ es el modelo que sólo tiene como predictor la constante.

Se observa fácilmente que entre las cuatro distribuciones utilizadas, la normal tiene el desvío más alto no importa si se considera el modelo con las interacciones entre tres factores, con interacciones simples, sin interacciones o de sólo la constante. Sin embargo, esto es sólo un indicador pues no podemos concluir nada salvo en el caso de la Poisson pues dado que se conoce $\sigma^2 = 1$, por (2.15) es posible comparar el desvío con una tabla $\chi^2$. En este caso, aceptamos el modelo pues $0.25 < 9.488 = \chi^2_{4,0.95}$ y dado que los modelos gama e inversa Gaussiana tienen un desvío menor también se pueden aceptar. Debido a que $5.33 << 9.49 = \chi^2_{(4)}$ entonces podemos aceptar también al modelo normal.

Es fácil observar que la diferencia de dos desvíos consecutivos cualesquiera coincide con la suma de cuadrados asociada al término que se retiró del modelo. Por ejemplo, la diferencia de desvíos entre el segundo y el tercer modelo es $12.000 - 6.833 = 5.167$, que es exactamente la suma de cuadrados asociada al término $CDE$ (el cual fue retirado del segundo modelo para obtener el tercero). Lo mismo ocurre con los grados de libertad. La diferencia de grados de libertad entre aquellos modelos es 2 que son los grados de libertad asociados a $CDE$.

TABLA 5. ANALISIS DE VARIANZA DEL EJEMPLO 1

| Fuente | Suma de Cuadrados | Grados de Libertad |
|---|---|---|
| B | 504.055 | 2 |
| C | 2949.055 | 2 |
| D | 2.250 | 1 |
| E | 240.250 | 1 |
| BC | 337.779 | 4 |
| BD | 5.167 | 2 |
| BE | 3.500 | 2 |
| CD | 4.167 | 2 |
| CE | 12.500 | 2 |
| DE | 2.250 | 1 |
| BCD | 8.667 | 4 |
| BCE | 10.000 | 4 |
| BDE | 1.500 | 2 |
| CDE | 5.168 | 2 |
| BCDE | 5.33 | 4 |
| Residual | 0 | 0 |
| Total | 4091.639 | 35 |

Debido a que sólo se tiene una observación por cada combinación de factores, la suma de cuadrados de los residuos es cero y no es posible estimar $\sigma^2$.

TABLA 6

ANALISIS DEL DESVIO - EJEMPLO 1

| Modelo | Desvío | g.l. |
|---|---|---|
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E+B.D.E+C.D.E+B.C.D.E$ | 0 | 0 |
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E+B.D.E+C.D.E$ | 5.333 | 4 |
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E+C.D.E$ | 6.833 | 6 |
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E$ | 12.000 | 8 |
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.E$ | 20.667 | 12 |
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E$ | 30.667 | 16 |
| $B+C+D+E+B.C+B.D+B.E+C.D+C.E$ | 32.917 | 17 |
| $B+C+D+E+B.C+B.D+C.D+C.E$ | 36.417 | 19 |
| $B+C+D+E+B.C+B.D+C.E$ | 40.583 | 21 |
| $B+C+D+E+B.C+C.E$ | 45.750 | 23 |
| $B+C+D+E+B.C$ | 58.250 | 25 |
| $B+C+D+E$ | 396.03 | 29 |
| $B+C+E$ | 398.28 | 30 |
| $B+C$ | 638.53 | 31 |
| $C$ | 1142.6 | 33 |
| Constante | 4091.6 | 35 |

TABLA 7. DESVIO DEL EJEMPLO 1

DIFERENTES DISTRIBUCIONES

| Modelo | Desvío | $f$ | $\Delta D_i$ | $\Delta f$ | $\hat{\sigma}^2$ | F |
|---|---|---|---|---|---|---|
| Normal | | | | | | |
| $M_1$ | 5.33 | 4 | | | 1.33 | |
| $M_2$ | 30.67 | 16 | 25.33 | 12 | 1.92 | 1.59 |
| $M_3$ | 396.03 | 29 | 365.36 | 13 | 13.66 | 14.64 |
| $M_4$ | 4091.60 | 35 | 3695.57 | 6 | 116.90 | 45.09 |
| Gamma | | | | | | |
| $M_1$ | 0.08 | 4 | | | 0.020 | |
| $M_2$ | 0.31 | 16 | 0.23 | 12 | 0.019 | 0.96 |
| $M_3$ | 1.82 | 29 | 1.51 | 13 | 0.063 | 6.11 |
| $M_4$ | 16.05 | 35 | 14.23 | 6 | 0.459 | 37.65 |
| Poisson | | | | | | |
| $M_1$ | 0.25 | 4 | | | | |
| $M_2$ | 2.43 | 16 | 2.18 | 12 | | |
| $M_3$ | 13.43 | 29 | 11.00 | 13 | | |
| $M_4$ | 238.52 | 35 | 225.11 | 6 | | |
| Inversa Gaussiana | | | | | | |
| $M_1$ | 0.0002 | 4 | | | 0.00005 | |
| $M_2$ | 0.006 | 16 | 0.0058 | 12 | 0.00038 | 9.67 |
| $M_3$ | 0.14 | 29 | 0.1340 | 13 | 0.00483 | 27.13 |
| $M_4$ | 1.30 | 35 | 1.16 | 6 | 0.03714 | 40.03 |

Dado que para un nivel de confianza del 95% se tienen $F_{12,4} = 5.91$; $F_{13,16} = 2.40$; $F_{6,29} = 2.43$ y que para un nivel de confianza del 99% se tienen $F_{12,4} = 14.37$; $F_{13,16} = 3.50$; $F_{6,29} = 3.50$ se observa fácilmente que en el caso normal el modelo $M_2$ (el que considera las interacciones) es el adecuado (si se toma un nivel de confianza del 99% y no del 95%) pues aunque $M_1$ también los ajusta, son innecesarias las interacciones de tres factores; en el caso gama el modelo $M_4$ más simple y que ajusta muy bien los datos y en el cuál además se observa que tanto los factores como sus interacciones son significativas mientras que las interacciones entre tres factores no es significativa; en el caso inversa Gaussiana (el mejor de todos pues los desvíos son muy pequeños) aunque el modelo $M_4$ es el mejor debido a su simplicidad, son significativas tanto los factores como las interacciones entre tres factores; en el caso Poisson se observa que el mejor modelo por su simplicidad es el $M_3$ siendo significativas las interacciones pero no así las interacciones entre tres factores.

Es posible analizar cada factor por separado en cada uno de estos modelos y obtener los estimados de los parámetros y de sus desviaciones estándar así como residuales sin mayor esfuerzo pues estamos dentro del contexto de los modelos lineales generalizados pero esto escaparía al objetivo principal del presente trabajo.

**Ejemplo 2.** Se generaron números aleatorios de distribución normal para que mediante SGPLUS de acuerdo al modelo siguiente :

$$y_{til} = A + B_t + C_i + (BC)_{ti} + \epsilon_{til}$$

donde $t = 1, 2, 3$; $i = 1, 2, 3, 4$; $l = 1, 2, 3$ y de manera que se cumplieran las condiciones $\sum B_t = 0$; $\sum C_i = 0$; $\sum (BC)_t = 0$; $\sum (BC)_i = 0$. Para ello se eligieron los siguientes valores para los parámetros : $A = 100$; $B_1 = 50$; $B_2 = -30$;

$B_3 = -20$; $C_1 = 16$; $C_2 = -10$; $C_3 = -12$; $C_4 = 6$; $(BC)_{11} = 2$; $(BC)_{12} = 1$; $(BC)_{13} = -1$; $(BC)_{14} = -2$; $(BC)_{21} = 3$; $(BC)_{22} = -3$; $(BC)_{23} = -1$; $(BC)_{24} = 1$; $(BC)_{31} = -5$; $(BC)_{32} = 2$; $(BC)_{33} = 2$; $(BC)_{14} = 1$.

Los datos obtenidos y el programa GLIM que se utilizó están en el apéndice. Los primera fila de datos corresponde a $y_{11l}$; la segunda a $y_{12l}; \cdots$. Se observa fácilmente que el modelo normal que ajusta es el que incluye el término de interacción mientras que el que sólo contiene los factores no ajusta (de acuerdo a como se generaron los datos). Sin embargo, los modelos gama y Gaussiana inversa ajustan mucho mejor aún considerando sólo el modelo constante. Por otro lado, no se incluye un ajuste con el modelo Poisson pues los números aleatorios generados son reales y no tendría sentido suponer una variable dependiente Poisson. En caso de multiplicar por 1000 las observaciones para poder utilizar una distribución Poisson, el análisis de desvío muestra que ninguno de los modelos sirve para explicar los datos.

Lo irónico de este ejemplo es que a pesar de generar una variable normal resultan ajustando mejor la distribución gama y Gaussiana inversa. Esto se debe a que no se genero una muestra grande (sólo 5 valores) en cada caso. Sin embargo, esto nos debe llevar a replantearnos seriamente la elección de la distribución normal como mecanismo que explica la parte aleatoria del modelo.

## Bibliografía


Cordeiro, G. (1986), *Modelos Lineares Generalizados VII SINAPE (Simposio Nacional de Probabilidade e Estatística)*, Universidades Estadual de Campinas, Campinas, São Paulo.

Dumett, M. (1995), *Modelos Lineales Generalizados (imprenta)*, Universidad Nacional de Colombia, Bogotá.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications Volume II Second Edition*, John Wiley, New York, N.Y..

Graybill, F. (1961), *An Introduction to Linear Statistical Models Volume I*, Mc Graw Hill, New York, N.Y..

Johnson, N. y Leone, F. (1964), *Statistics and Experimental Design*, John Wiley, New York, N.Y..

Jorgensen, B. (1992), *The Theory of Exponencial Dispersion Models and Analysis of Deviance (Monografias de Matemáticas N 51)*, IMPA (Instituto de Matemática Pura e Aplicada, Rio

de Janeiro.
Scheffé, H. (1959), *The Analysis of Variance*, John Wiley, New York, N.Y..

## APENDICE

A CONTINUACION SE PRESENTA EL PROGRAMA GLIM UTILIZADO EN EL EJEMPLO 1

```
[o] GLIM 3.77 update 1 (copyright)1985 Royal Statistical Society, London
[o]
[i] ? $TRA I O$
[i] ? $INPUT 7 80 STAM$
[i] File name? DEZVIO.GLM
[i] $UNITS 36$DATA MZ$READ
[i] 8 5 8 4 17 11 13 10 22 16 20 15 7 3 10 5 26 17 24 19 34 32 34 29 10 5 9 4
[i] 24 14 24 16 39 33 36 34
[i] $FAC B 3 C 3 D 2 E 2$YVAR MZ $
[i] $CAL B=
[i] $ERR N$FIT B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E
+B.D.E+C.D.E :
[o] deviance = 5.3333
[o] d.f. = 4
[o]
[i] -B.C.D-B.C.E-B.D.E-C.D.E : -B.C-B.D-B.E-C.D-C.E-D.E : -B-C-D-E$
[o] deviance = 30.667 (change = +25.33)
```

[o] d.f. = 16 (change = +12 )

[o]

[o] deviance = 396.03 (change = +365.4)

[o] d.f. = 29 (change = +13 )

[o]

[o] deviance = 4091.6 (change = +3696.)

[o] d.f. = 35 (change = +6 )

[o]

[i] $ERR G$FIT B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E

+B.D.E+C.D.E :

[o] deviance = 0.080538 at cycle 3

[o] d.f. = 4

[o]

[i] -B.C.D-B.C.E-B.D.E-C.D.E : -B.C-B.D-B.E-C.D-C.E-D.E : -B-C-D-E$

[o] deviance = 0.31373 (change = +0.2332) at cycle 3

[o] d.f. = 16 (change = +12 )

[o]

[o] deviance = 1.8243 (change = +1.511) at cycle 4

[o] d.f. = 29 (change = +13 )

[o]

[o] deviance = 16.054 (change = +14.23) at cycle 4

[o] d.f. = 35 (change = +6 )

[o]

[i] $ERR P$FIT B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E

```
+B.D.E+C.D.E :
[o] scaled deviance = 0.25084 at cycle 3
[o] d.f. = 4
[o]
[i] -B.C.D-B.C.E-B.D.E-C.D.E : -B.C-B.D-B.E-C.D-C.E-D.E : -B-C-D-E$
[o] scaled deviance = 2.4294 (change = +2.179) at cycle 3
[o] d.f. = 16 (change = +12 )
[o]
[o] scaled deviance = 13.425 (change = +11.00) at cycle 3
[o] d.f. = 29 (change = +13 )
[o]
[o] scaled deviance = 238.52 (change = +225.1) at cycle 4
[o] d.f. = 35 (change = +6 )
[o]
[i] $MACRO ENLA $CAL
[i] $MACRO DERI $CAL
[i] $MACRO VARI $CAL
[i] $MACRO DVIO $CAL
[i] $OWN ENLA DERI VARI DVIO $
[i] $CAL
[i] $FIT B+C+D+E+B.C+B.D+B.E+C.D+C.E+D.E+B.C.D+B.C.E+B.D.E
+C.D.E :
[o] deviance = 0.00026049 at cycle 3
[o] d.f. = 4
```

[o]

[i] -B.C.D-B.C.E-B.D.E-C.D.E : -B.C-B.D-B.E-C.D-C.E-D.E : -B-C-D-E$

[o] deviance = 0.0062383 (change = +0.005978) at cycle 3

[o] d.f. = 16 (change = +12 )

[o]

[o] deviance = 0.14012 (change = +0.1339) at cycle 4

[o] d.f. = 29 (change = +13 )

[o]

[o] deviance = 1.301 (change = +1.16) at cycle 3

[o] d.f. = 35 (change = +6 )

[o]

[i] $RETURN$

[i] ? $STOP$

A CONTINUACION SE PRESENTA EL PROGRAMA GLIM UTILIZADO EN EL EJEMPLO 2

[o] GLIM 3.77 update 1 (copyright)1985 Royal Statistical Society, London

[o]

[i] ? $TRA I O$

[i] ? $INPUT 7 80 STAM$

[i] File name? RANDOM.GLM

[i] $UNITS 60$DATA ART$READ

[i] 168.479 166.494 167.623 169.410 167.879

[i] 139.595 140.619 141.866 141.226 140.627

```
[i] 136.306 137.169 136.926 136.129 137.083
[i] 154.810 152.415 153.380 154.528 154.012
[i] 88.068 89.646 88.395 89.829 90.394
[i] 58.145 57.038 58.179 57.376 57.030
[i] 56.683 57.108 55.986 57.691 57.313
[i] 76.523 77.716 76.725 77.596 77.254
[i] 89.908 92.444 91.307 91.909 91.057
[i] 82.720 81.224 82.317 82.852 80.843
[i] 69.665 70.866 69.168 69.346 69.401
[i] 86.729 88.391 87.918 85.378 85.394 $
[i] $FAC B 3 C 4$YVAR ART $
[i] $CAL B=
[i] $ERR N$FIT B+C+B.C : -B.C : -B : -C $
[o] deviance = 36.697
[o] d.f. = 48
[o]
[o] deviance = 773.63 (change = +736.9)
[o] d.f. = 54 (change = +6 )
[o]
[o] deviance = 74275. (change = +73502.)
[o] d.f. = 56 (change = +2 )
[o]
[o] deviance = 81571. (change = +7296.)
[o] d.f. = 59 (change = +3 )
```

[o]

[i] $ERR G$FIT $

[o] deviance = 7.7663 at cycle 4

[o] d.f. = 59

[o]

[i] $MACRO ENLA $CAL

[i] $MACRO DERI $CAL

[i] $MACRO VARI $CAL

[i] $MACRO DVIO $CAL

[i] $OWN ENLA DERI VARI DVIO $

[i] $CAL

[i] $FIT $

[o] deviance = 0.07996 at cycle 3

[o] d.f. = 59

[o]

[i] $RETURN$

[i] ? $STOP$

Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia, BOGOTA.