

MODELOS LINEALES GENERALIZADOS MIXTOS ALGUNOS CASOS PRÁCTICOS

Llorenç Badiella
Servei d'Estadística Aplicada
Universitat Autònoma de Barcelona

X Congreso Galego de Estatística e Investigación de Operacións – Pontevedra

3–4–5 de novembro de 2011

Contenidos

1: Introducción

2: Modelos Lineales Mixtos

- Factores Aleatorios (Ansiolíticos)
- Modelos para pendientes aleatorias (Pino Rojo)
- Modelos para estructuras de covarianzas (Ansiolíticos)

3: Modelos Mixtos para datos no normales

- Modelo Condicional: GLMM (Brotos clínicos)
- Modelo Marginal: GEE (Taekwondo)

Introducción

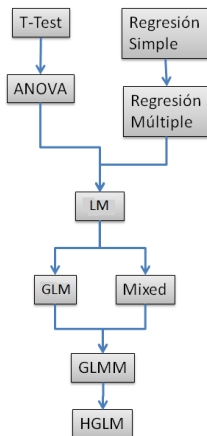
Los problemas estadísticos son de dos clases:

- Descubrir qué cantidades son necesarias para la descripción adecuada de una población, hecho que incluye la consideración de expresiones matemáticas para representar las distribuciones de frecuencia.
- Determinar cuanta información y de qué tipo respecto a tales valores poblacionales se refleja en una o varias muestras aleatorias.

R. Fisher (1921)

Introducción

Clasificación de modelos



Introducción

Definición Modelo lineal generalizado (GLM)

Propuesta de análisis adecuada cuando las observaciones son independientes pero las suposiciones sobre la Normalidad de la variable respuesta (y los errores) son incumplidas.

Definición Modelo mixto (MIXED)

Propuesta de modelo estadístico adecuada cuando la variable respuesta sigue una distribución Normal, permitiendo modelizar su variabilidad y la presencia de observaciones correlacionadas.

Introducción

Definición Modelo lineal generalizado (GLM)

Propuesta de análisis adecuada cuando las observaciones son independientes pero las suposiciones sobre la Normalidad de la variable respuesta (y los errores) son incumplidas.

Definición Modelo mixto (MIXED)

Propuesta de modelo estadístico adecuada cuando la variable respuesta sigue una distribución Normal, permitiendo modelizar su variabilidad y la presencia de observaciones correlacionadas.

Introducción

Nomenclatura para modelos lineales generalizados

- Regresión Logística
- Regresión de Poisson
- Modelo ZIP
- Regresión Binomial Negativa

Nomenclatura para modelos mixtos

- Modelos multinivel (sociología)
- Modelos jerárquicos o anidados (psicología)
- Meta-análisis (medicina)
- Modelos para datos longitudinales (medicina)
- Modelos para medidas repetidas (biología)
- Modelos con coeficientes aleatorios (estadística)

Introducción

Nomenclatura para modelos lineales generalizados

- Regresión Logística
- Regresión de Poisson
- Modelo ZIP
- Regresión Binomial Negativa

Nomenclatura para modelos mixtos

- Modelos multinivel (sociología)
- Modelos jerárquicos o anidados (psicología)
- Meta-análisis (medicina)
- Modelos para datos longitudinales (medicina)
- Modelos para medidas repetidas (biología)
- Modelos con coeficientes aleatorios (estadística)

Introducción

Definición Modelo lineal generalizado mixto (GLMM)

Los modelos lineales generalizados mixtos constituyen una fusión entre el modelo lineal mixto y el modelo lineal generalizado.

A continuación se presentan las propuestas de modelización más comunes empleando modelos GLMM:

- Modelo Mixto con un sólo factor aleatorio (Ansiolíticos)
- Modelo Mixto para pendientes aleatorias (Pino Rojo)
- Modelo Mixto para estructuras de covarianzas (Ansiolíticos)
- Modelo GLMM Marginal (Taekwondo)
- Modelo GLMM Condicional (Brotos clínicos)

Introducción

Definición Modelo lineal generalizado mixto (GLMM)

Los modelos lineales generalizados mixtos constituyen una fusión entre el modelo lineal mixto y el modelo lineal generalizado.

A continuación se presentan las propuestas de modelización más comunes empleando modelos GLMM:

- Modelo Mixto con un sólo factor aleatorio (Ansiolíticos)
- Modelo Mixto para pendientes aleatorias (Pino Rojo)
- Modelo Mixto para estructuras de covarianzas (Ansiolíticos)
- Modelo GLMM Marginal (Taekwondo)
- Modelo GLMM Condicional (Brotos clínicos)

Modelos Lineales Mixtos

Un modelo lineal mixto es una generalización del modelo lineal general contemplando:

- Variable respuesta con distribución Normal
- Efectos fijos y efectos aleatorios (efectos mixtos)
- Posibilidad de datos correlacionados
- Variabilidad heterogénea

Modelo para un sólo factor aleatorio

El modelo lineal con un sólo factor aleatorio responde al siguiente planteamiento:

Modelo

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \dots + \beta_p X_{pij} + u_i + \epsilon_{ij}$$

- Se han seleccionado una serie de niveles u_i de una determinada población heterogénea de clústers:
 $u_i \sim N(0, \sigma_U^2)$.
- En cada uno de estos niveles preseleccionados u_i se han realizado una serie de mediciones, con cierto error experimental: $\epsilon_{ij} \sim N(0, \sigma^2)$.

Modelo para un sólo factor aleatorio

El modelo lineal con un sólo factor aleatorio responde al siguiente planteamiento:

Modelo

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \dots + \beta_p X_{pij} + u_i + \epsilon_{ij}$$

- Se han seleccionado una serie de niveles u_i de una determinada población heterogénea de clústers:
 $u_i \sim N(0, \sigma_U^2)$.
- En cada uno de estos niveles preseleccionados u_i se han realizado una serie de mediciones, con cierto error experimental: $\epsilon_{ij} \sim N(0, \sigma^2)$.

Modelo para un sólo factor aleatorio

En particular este planteamiento genera de forma automática una estructura de correlaciones:

Propiedades

- $E[Y_{ij}] = \beta_0 + \beta_1 X_{1ij} + \dots + \beta_p X_{pij}$
- $\text{Var}[Y_{ij}] = \sigma_U^2 + \sigma^2$
- $\text{Cov}[Y_{ij}, Y_{i'j'}] = 0$
- $\text{Cov}[Y_{ij}, Y_{ij'}] = \sigma_U^2$

Ejemplo: Ansiolítico

- 10 individuos son sometidos a pruebas para medir ansiedad.
- 4 periodos diferentes, evaluando 4 dosis de ansiolítico.
- En cada periodo, condición control y ansiedad provocada.
- La variable respuesta es en una medida eléctrica del sobresalto producido por ciertos estímulos.

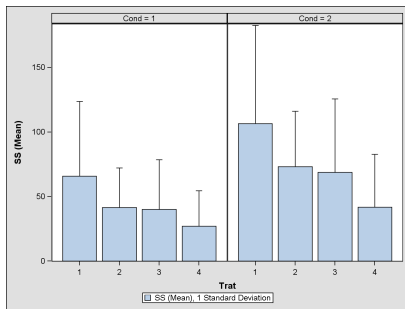


Figura: Estudio Ansiolíticos

Modelo de *intercepts* y *slopes* aleatorios

Este modelo responde al siguiente planeamiento:

Modelo

$$Y_{ij} = \beta_0 + u_i + (\beta_1 + v_i)X_{1ij} + \epsilon_{ij}$$

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + (u_i + v_i X_{1ij}) + \epsilon_{ij}$$

Modelo de *intercepts* y *slopes* aleatorios

Este modelo responde al siguiente planeamiento:

Modelo

$$Y_{ij} = \beta_0 + u_i + (\beta_1 + v_i)X_{1ij} + \epsilon_{ij}$$

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + (u_i + v_i X_{1ij}) + \epsilon_{ij}$$

- Es adecuado para analizar estudios con datos longitudinales desestructurados.
- Se asume que en el término independiente y en el pendiente de cada individuo se añade un ruido aleatorio: $u_i \sim N(0, \sigma_U^2)$, $v_i \sim N(0, \sigma_V^2)$.
- Existe un error residual: $\epsilon_{ij} \sim N(0, \sigma^2)$
- Se genera automáticamente una estructura de correlaciones gradual según la proximidad de las medidas.

Modelo de *intercepts* y *slopes* aleatorios

Este modelo responde al siguiente planeamiento:

Modelo

$$Y_{ij} = \beta_0 + u_i + (\beta_1 + v_i)X_{1ij} + \epsilon_{ij}$$

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + (u_i + v_i X_{1ij}) + \epsilon_{ij}$$

- Es adecuado para analizar estudios con datos longitudinales desestructurados.
- Se asume que en el término independiente y en el pendiente de cada individuo se añade un ruido aleatorio: $u_i \sim N(0, \sigma_U^2)$, $v_i \sim N(0, \sigma_V^2)$.
- Existe un error residual: $\epsilon_{ij} \sim N(0, \sigma^2)$
- Se genera automáticamente una estructura de correlaciones gradual según la proximidad de las medidas.

Ejemplo: Pino Rojo

- Se pretende comparar el crecimiento del pino rojo según el período de nacimiento.
- Se incluyen 35 parcelas. En cada parcela se escogieron entre 3 y 7 árboles.
- Las fechas de nacimiento fueron agrupadas en 4 periodos: 1900-1925, 1926-1950, 1951-1975, 1976-1990.

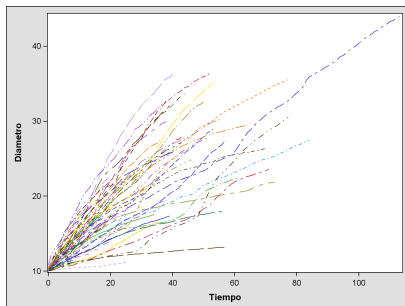


Figura: Estudio Pino Rojo

Modelos de estructuras de correlaciones

Los modelos anteriores pueden escribirse de forma general empleando notación matricial. Por ejemplo:

$$Y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + u_i + \epsilon_{ij}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \\ \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_q \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \dots \\ \epsilon_n \end{pmatrix}$$

Dando lugar a la formulación matricial.

$$Y = X\beta + Zu + \epsilon$$

Modelos de estructuras de correlaciones

Los modelos anteriores pueden escribirse de forma general empleando notación matricial. Por ejemplo:

$$Y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + u_i + \epsilon_{ij}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \\ \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_q \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \dots \\ \epsilon_n \end{pmatrix}$$

Dando lugar a la formulación matricial.

$$Y = X\beta + Zu + \epsilon$$

Modelos de estructuras de correlaciones

De esta manera:

Formulación matricial

$$Y = X\beta + Zu + \epsilon$$

$$\epsilon \sim N(0, I_n\sigma^2)$$

$$u \sim N(0, I_q\sigma_u^2)$$

$$Y|u \sim N(X\beta + Zu, I_n\sigma^2)$$

$$Y \sim N(X\beta, ZGZ' + \sigma^2I_n)$$

Finalmente, de forma genérica:

$$Y \sim N(X\beta, R)$$

Modelos de estructuras de correlaciones

Los modelos mixtos planteados de forma marginal y empleando notación matricial corresponden a:

$$Y \sim N(X\beta, R)$$

donde R es una matriz normalmente definida por bloques (niveles o individuos).

Distintas propuestas de estructura son:

- Sin estructura,
- Con simetría compuesta,
- AR(1),
- versiones heterogéneas, etc.

Modelos de estructuras de correlaciones

Simetría compuesta

Estructura de covarianzas por bloque:

$$B_V = \begin{pmatrix} \sigma^2 + \sigma_U^2 & \sigma_U^2 & \sigma_U^2 & \sigma_U^2 \\ \sigma_U^2 & \sigma^2 + \sigma_U^2 & \sigma_U^2 & \sigma_U^2 \\ \sigma_U^2 & \sigma_U^2 & \sigma^2 + \sigma_U^2 & \sigma_U^2 \\ \sigma_U^2 & \sigma_U^2 & \sigma_U^2 & \sigma^2 + \sigma_U^2 \end{pmatrix}$$

Estructura de correlaciones por bloque:

$$B_R = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

Modelos de estructuras de correlaciones

Autorregresiva

Estructura de covarianzas por bloque:

$$B_V = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

Estructura de correlaciones por bloque:

$$B_R = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

Modelos de estructuras de correlaciones

Estructura general (sin estructura)

Estructura de covarianzas por bloque:

$$B_V = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \sigma_{23}^2 & \sigma_{24}^2 \\ \sigma_{13}^2 & \sigma_{23}^2 & \sigma_{33}^2 & \sigma_{34}^2 \\ \sigma_{14}^2 & \sigma_{24}^2 & \sigma_{34}^2 & \sigma_{44}^2 \end{pmatrix}$$

Modelos de estructuras de correlaciones

De esta manera, existen dos versiones diferentes del mismo planteamiento:

Modelo de efectos aleatorios (modelo tipo G)

$$Y \sim N(X\beta, ZGZ' + I_n\sigma^2)$$

Modelo de estructura de correlaciones (modelo tipo R)

$$Y \sim N(X\beta, R)$$

La segunda propuesta es más general e incluye la primera, no obstante, cuando existe correspondencia, ambos modelos son el mismo.

Modelos lineales generalizados

En numerosas ocasiones la variable respuesta no sigue una distribución Normal: una variable discreta, una tasa, una proporción, ... En tal caso,

- Las relaciones lineales no son válidas (predicciones fuera de rango).
- Los errores no suelen ser homogéneos.
- La inferencia basada en modelos lineales no suele ser adecuada.
- Aplicando transformaciones (log, Sqrt, ...) el modelo se distorsiona
- Es necesario emplear modelos que se adapten a la propia distribución de los errores: Modelos GLM o GLMM.

Modelos lineales generalizados

Especificación del Modelo GLM

Modelo

$$E[Y] = \mu$$

$$g(\mu) = X\beta$$

$$\text{Var}[Y] = V(\mu)\phi$$

- μ es el valor esperado de la variable objetivo.
- $g()$ es la función de enlace (link).
- $V(\mu)$ es la función varianza asociada a la distribución de los datos.
- ϕ es un parámetro de escala y permite recoger una posible sobredispersión.

Modelos lineales generalizados

Especificación del Modelo GLM

Modelo

$$E[Y] = \mu$$

$$g(\mu) = X\beta$$

$$\text{Var}[Y] = V(\mu)\phi$$

- μ es el valor esperado de la variable objetivo.
- $g()$ es la función de enlace (link).
- $V(\mu)$ es la función varianza asociada a la distribución de los datos.
- ϕ es un parámetro de escala y permite recoger una posible sobredispersión.

Modelo Condicional: GLMM

El modelo condicional (G) es el siguiente:

Modelo Condicional

$$E[Y|u] = g^{-1}(X\beta + Zu)$$
$$u \sim N(0, G)$$

- Se asume que condicional a los efectos aleatorios se trata de un modelo lineal generalizado simple y que la distribución de los efectos aleatorios u es Normal.
- La introducción de efectos aleatorios en el modelo, induce una estructura de correlaciones en la distribución marginal.
- El método de estimación emplea (aproxima) la función de verosimilitud.

Modelo Condicional: GLMM

El modelo condicional (G) es el siguiente:

Modelo Condicional

$$E[Y|u] = g^{-1}(X\beta + Zu)$$
$$u \sim N(0, G)$$

- Se asume que condicional a los efectos aleatorios se trata de un modelo lineal generalizado simple y que la distribución de los efectos aleatorios u es Normal.
- La introducción de efectos aleatorios en el modelo, induce una estructura de correlaciones en la distribución marginal.
- El método de estimación emplea (aproxima) la función de verosimilitud.

Ejemplo: Recuentos longitudinales-Brotos clínicos

- Se desea analizar el patrón de evolución de una patología auto-inmune (número de brotes)
- La información de los pacientes es recogida semestralmente.
- Se incluyen 277 pacientes con 3 o más visitas.
- Se desea analizar la influencia de la edad en el momento de diagnóstico o del sexo en la severidad de la enfermedad.

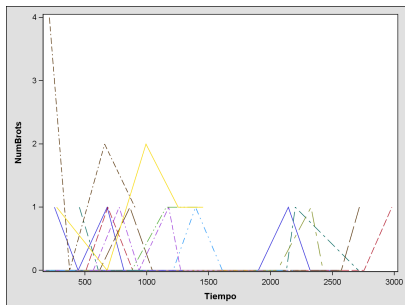


Figura: Estudio Brotes Clínicos

Modelo Marginal: GEE

El modelo marginal (R) se concreta del modo siguiente:

Modelo Marginal

$$g(E[Y]) = X\beta$$

$$\text{Var}[Y] = V(\mu)\phi$$

$$\text{Corr}[Y_{ij}, Y_{ij'}] = \alpha_{jj'}$$

- La estructura de correlaciones se define a partir de los parámetros $\alpha_{jj'}$.
- El método de estimación emplea una función de cuasi-verosimilitud: las ecuaciones de estimación generales (generalized estimating equations (GEE)).

Modelo Marginal: GEE

El modelo marginal (R) se concreta del modo siguiente:

Modelo Marginal

$$g(E[Y]) = X\beta$$

$$\text{Var}[Y] = V(\mu)\phi$$

$$\text{Corr}[Y_{ij}, Y_{ij'}] = \alpha_{jj'}$$

- La estructura de correlaciones se define a partir de los parámetros $\alpha_{jj'}$.
- El método de estimación emplea una función de quasi-verosimilitud: las ecuaciones de estimación generales (generalized estimating equations (GEE)).

Ejemplo: GLMM Recuentos - Combates Taekwondo

- Se desea determinar qué factores influyen en la cantidad y éxito de movimientos ofensivos realizados por los contrincantes.
- 32 combates diferentes pertenecientes a distintas categorías de peso, y en total, 2246 movimientos ofensivos.
- Los datos han sido agrupados según intervalos de tiempo de 1 minuto de duración.

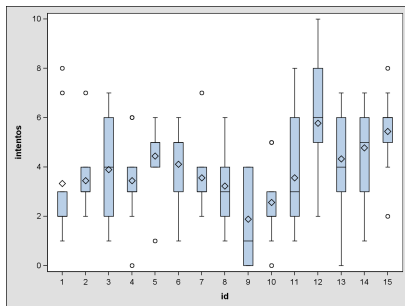


Figura: Combates Taekwondo

Diferencias entre modelo Condicional y Marginal

Perspectiva técnica

- Empleando GEE las componentes de la varianza se estiman por el método de los momentos.

Perspectiva interpretativa

- En GLMM los coeficientes representan el efecto a nivel individual, mientras que en GEE en el promedio poblacional.
- Empleando GEE la estructura de correlaciones es tratada aparte.

Modelos GLMM y software

V. Respuesta	Observaciones	SAS	R
Normal	Indep.	GLM	lm
Normal	Corr.	MIXED	lme(nlme) lmer(lme4)
F. Expo.	Indep.	GENMOD	glm
F. Expo.	Corr. Cond.	GLIMMIX	glmer(lme4)
F. Expo.	Corr. Marginal	GLIMMIX	geeglm(lme4)

Referencias

- Badiella, L. and Pedromingo, A. (2006). *Métodos multivariantes en biomedicina. Casos prácticos y fundamentos teóricos con el programa G-Stat 2.0*. Fundación de Ciencias de la Salud.
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects*, Chapman & Hall/CRC Boca Raton.
- Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for Mixed Models*, SAS Institute Inc..
- McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, Second Ed. Chapman and Hall/CRC, London.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*, New York: Springer.
- Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Zuur, A.F. et al. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Moitas Gracias!

En cualquier proyecto de investigación,
la **Estadística** juega un papel central
para que la información se convierta en
conocimiento

Asesórese desde el principio

Servei d'Estadística

Asesoramiento y consultoría experta en todos los ámbitos de la Estadística

Servei d'Estadística
Universitat Autònoma de Barcelona

Modelos lineales generalizados mixtos.

Algunos casos prácticos.

Llorenç BADIELLA
Servei d'Estadística Aplicada
Universitat Autònoma de Barcelona

X Congreso SGAPEIO

5 de Noviembre de 2011

Índice general

1. Algunos casos prácticos	5
1.1. Ansiolíticos	6
1.2. Pino Rojo	16
1.3. Ansiolíticos (2)	31
1.4. Episodios Enfermedad	42
1.5. Deporte de lucha	50

Capítulo 1

Algunos casos prácticos

1.1. Ansiolíticos

Este ejemplo permite mostrar el modelo mixto con un factor aleatorio donde el factor aleatorio genera bloques de subréplicas: en cada una de las medidas repetidas realizadas en un mismo nivel, las condiciones experimentales son distintas.

Un investigador ha seleccionado 10 individuos, sometidos a distintas pruebas para medir su ansiedad. En cuatro periodos diferentes se evalúan 4 dosis de un mismo tratamiento ansiolítico. En cada periodo, el efecto del tratamiento es medido en dos condiciones diferentes: una situación control y una situación de ansiedad provocada. La variable respuesta consiste en una medida eléctrica del sobresalto producido por ciertos (*startle response*) estímulos. El investigador desea evaluar el efecto de la dosis en ambas condiciones experimentales.

A continuación se presenta una muestra de los datos recogidos:

```
proc print data=a(obs=10);  
class Trat Cond;  
var ss;  
run;
```

Obs	Vol	Trat	Cond	SS
1	1	1	1	87.89
2	1	1	2	154.94
3	1	2	1	23.01
4	1	2	2	35.30
5	1	3	1	39.84
6	1	3	2	74.81
7	1	4	1	14.60
8	1	4	2	19.40
9	2	1	1	95.72
10	2	1	2	129.70

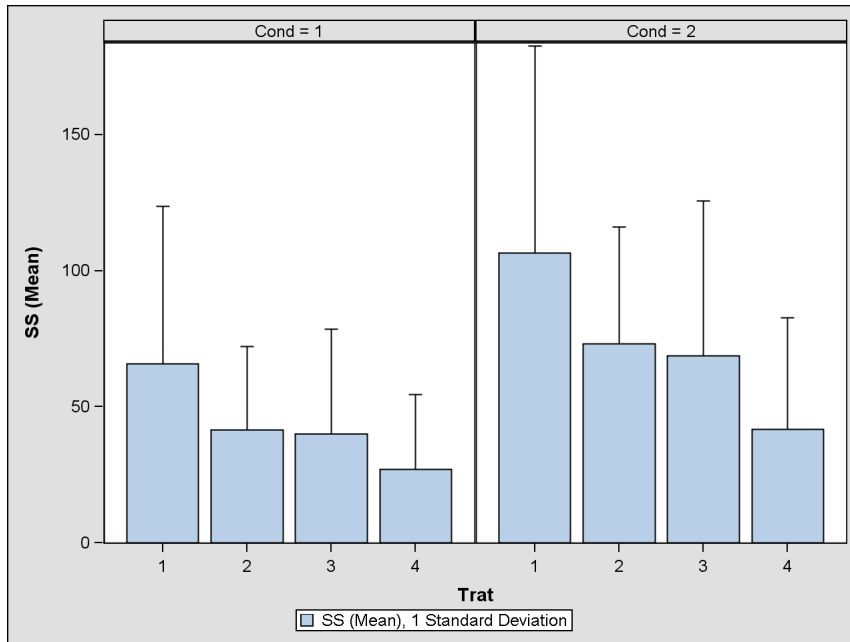
El resumen de los datos es el siguiente:

```
proc means data=a maxdec=2;  
class Trat Cond;  
var ss;  
run;
```

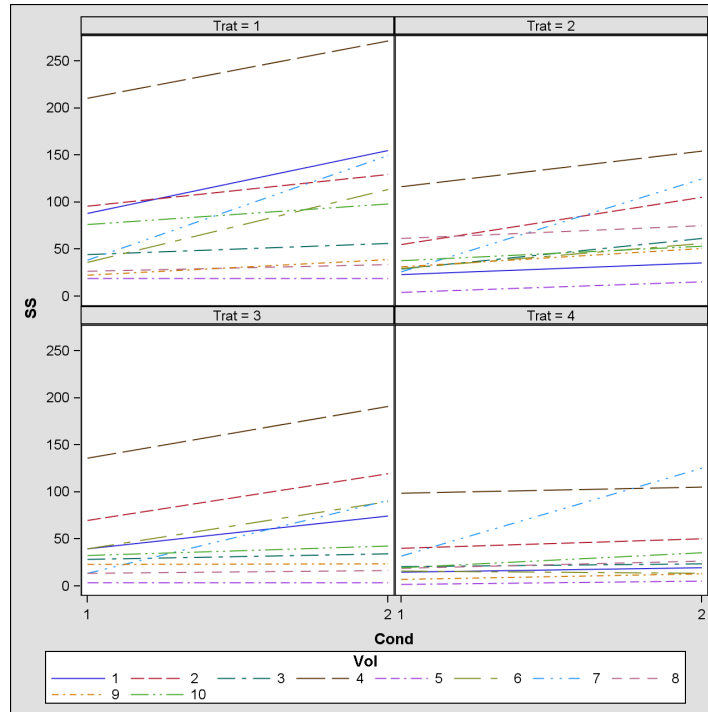
Analysis Variable : SS							
Trat	Cond	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	1	10	10	65.71	57.90	18.89	210.64
	2	10	10	106.50	76.12	19.02	271.48
2	1	10	10	41.39	30.88	4.16	116.30
	2	10	10	73.25	42.74	15.11	154.61
3	1	10	10	40.00	38.43	3.47	135.88
	2	10	10	68.67	57.09	3.36	190.76
4	1	10	10	26.95	27.49	1.68	98.57
	2	10	10	41.84	40.94	5.63	125.44

Gráficamente:

```
proc sgpanel data=a;  
panelby cond ;  
vbar trat / response=ss stat=mean limits=upper limitstat=stddev;  
run;
```




```
proc sgpanel data=a;  
panelby trat ;  
series x=cond2 y=ss / group=vol ;  
run;
```



El modelo para analizar el estudio, debería incluir un único factor aleatorio: el sujeto, identificado por la variable VOL. De esta manera, se tendrá en cuenta que existen diferentes medidas repetidas para cada sujeto, y las diferencias entre tratamientos y condiciones serán evaluadas a nivel intra-individual.

```
proc mixed data=a;
class trat vol cond;
model ss = trat cond trat*cond / dfm=sat residual solution;
random intercept / subject=vol v=1 vcorr=1;
lsmeans trat / adjust=tukey;
run;
```

Los índices de ajuste:

Fit Statistics	
-2 Res Log Likelihood	721.4
AIC (smaller is better)	725.4
AICC (smaller is better)	725.6
BIC (smaller is better)	726.0

Las componentes de la varianza:

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	Vol	1688.24
Residual		698.58

La estructura de varianzas y correlaciones resultante es la siguiente:

Estimated V Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	2386.82	1688.24	1688.24	1688.24	1688.24	1688.24	1688.24	1688.24
2	1688.24	2386.82	1688.24	1688.24	1688.24	1688.24	1688.24	1688.24
3	1688.24	1688.24	2386.82	1688.24	1688.24	1688.24	1688.24	1688.24
4	1688.24	1688.24	1688.24	2386.82	1688.24	1688.24	1688.24	1688.24
5	1688.24	1688.24	1688.24	1688.24	2386.82	1688.24	1688.24	1688.24
6	1688.24	1688.24	1688.24	1688.24	1688.24	2386.82	1688.24	1688.24
7	1688.24	1688.24	1688.24	1688.24	1688.24	1688.24	2386.82	1688.24
8	1688.24	1688.24	1688.24	1688.24	1688.24	1688.24	1688.24	2386.82

Estimated V Correlation Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.7073	0.7073	0.7073	0.7073	0.7073	0.7073	0.7073
2	0.7073	1.0000	0.7073	0.7073	0.7073	0.7073	0.7073	0.7073
3	0.7073	0.7073	1.0000	0.7073	0.7073	0.7073	0.7073	0.7073
4	0.7073	0.7073	0.7073	1.0000	0.7073	0.7073	0.7073	0.7073
5	0.7073	0.7073	0.7073	0.7073	1.0000	0.7073	0.7073	0.7073
6	0.7073	0.7073	0.7073	0.7073	0.7073	1.0000	0.7073	0.7073
7	0.7073	0.7073	0.7073	0.7073	0.7073	0.7073	1.0000	0.7073
8	0.7073	0.7073	0.7073	0.7073	0.7073	0.7073	0.7073	1.0000

A continuación, la tabla ANOVA para los efectos fijos y las estimaciones de los parámetros:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Trat	3	63	12.99	<.0001
Cond	1	63	24.17	<.0001
Trat*Cond	3	63	0.83	0.4844

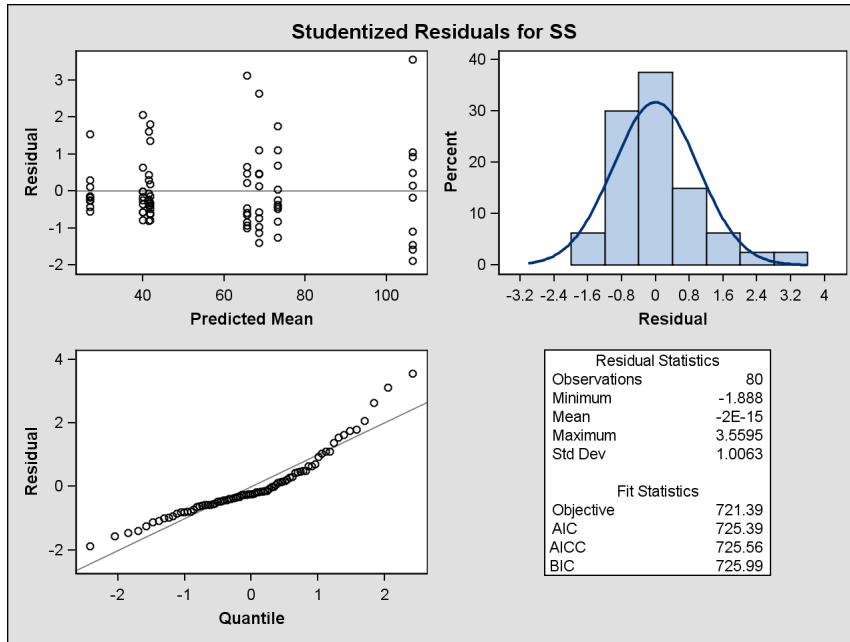
Solution for Fixed Effects							
Effect	Trat	Cond	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			41.8400	15.4493	16	2.71	0.0155
Trat	1		64.6630	11.8201	63	5.47	<.0001
Trat	2		31.4080	11.8201	63	2.66	0.0100
Trat	3		26.8300	11.8201	63	2.27	0.0266
Trat	4		0
Cond		1	-14.8940	11.8201	63	-1.26	0.2123
Cond		2	0
Trat*Cond	1	1	-25.8990	16.7162	63	-1.55	0.1263
Trat*Cond	1	2	0
Trat*Cond	2	1	-16.9620	16.7162	63	-1.01	0.3141
Trat*Cond	2	2	0
Trat*Cond	3	1	-13.7780	16.7162	63	-0.82	0.4129
Trat*Cond	3	2	0
Trat*Cond	4	1	0
Trat*Cond	4	2	0

Finalmente, la tabla de medias ajustadas LSMEANS y las comparaciones a posteriori aplicando la corrección de Tukey por multiplicidad de contrastes:

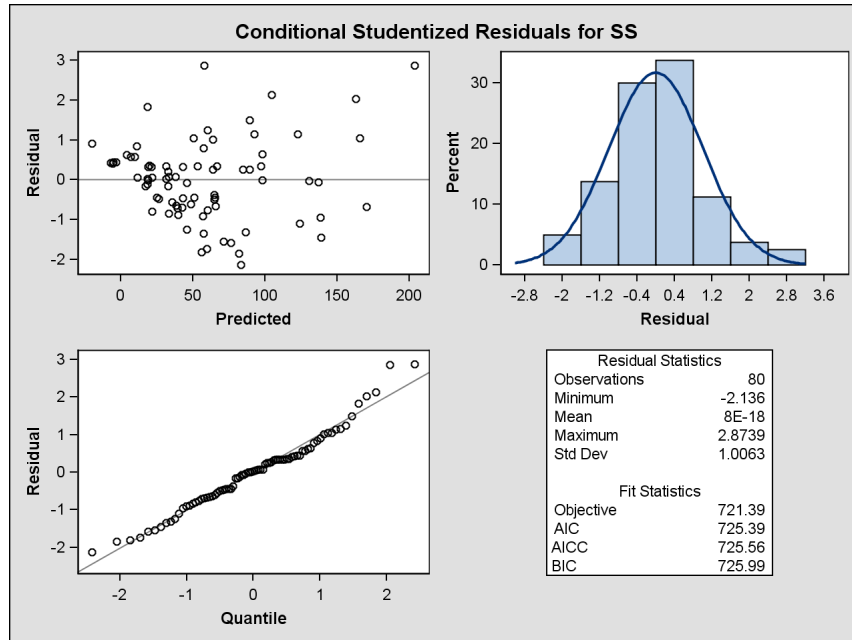
Least Squares Means						
Effect	Trat	Estimate	Standard Error	DF	t Value	Pr > t
Trat	1	86.1065	14.2742	11.8	6.03	<.0001
Trat	2	57.3200	14.2742	11.8	4.02	0.0018
Trat	3	54.3340	14.2742	11.8	3.81	0.0026
Trat	4	34.3930	14.2742	11.8	2.41	0.0332

Differences of Least Squares Means									
Effect	Trat	_Trat	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
Trat	1	2	28.7865	8.3581	63	3.44	0.0010	Tukey–Kramer	0.0055
Trat	1	3	31.7725	8.3581	63	3.80	0.0003	Tukey–Kramer	0.0018
Trat	1	4	51.7135	8.3581	63	6.19	<.0001	Tukey–Kramer	<.0001
Trat	2	3	2.9860	8.3581	63	0.36	0.7221	Tukey–Kramer	0.9842
Trat	2	4	22.9270	8.3581	63	2.74	0.0079	Tukey–Kramer	0.0386
Trat	3	4	19.9410	8.3581	63	2.39	0.0201	Tukey–Kramer	0.0903

Con el objetivo de validar las suposiciones de Normalidad de los residuos, podemos obtener el diagrama de residuos marginales estudentizados:



o los residuos condicionales studentizados:



En los diagramas de residuos anteriores se observa cierta falta de homogeneidad en la variabilidad de los residuos. De hecho, en el resumen descriptivo se observó una mayor variabilidad en los valores elevados de la variable respuesta. Este comportamiento es natural por el hecho de tratarse de una variable de escala. Sería conveniente rehacer el análisis empleando la transformación *logarítmica* en los datos.

1.2. Pino Rojo

El presente estudio pretende comparar el crecimiento del pino rojo según diferentes períodos. 35 parcelas fueron seleccionadas, y en cada parcela se escogieron un número de árboles cuyas fechas de nacimiento pertenecieran a los intervalos de tiempo 1900-1925, 1926-1950, 1951-1975, 1976-1990. El nacimiento se determinó a partir del momento en que el árbol medía 10 cm de diámetro. La variable Y4 identifica el periodo de nacimiento con los valores 0, 1, 2 y 3 respectivamente. El crecimiento fue evaluado examinando la amplitud de los diferentes anillos.

A continuación se presenta una muestra de los datos recogidos:

```
proc print data=a(obs=10);  
var Parcela Id Y4 Year Diametro;  
run;
```

Obs	parcela	id	y4	year	Diametro
1	BF628	1	2	1953	10.098
2	BF628	1	2	1954	10.262
3	BF628	1	2	1955	10.474
4	BF628	1	2	1956	10.674
5	BF628	1	2	1957	10.928
6	BF628	1	2	1958	11.190
7	BF628	1	2	1959	11.432
8	BF628	1	2	1960	11.622
9	BF628	1	2	1961	11.858
10	BF628	1	2	1962	12.076

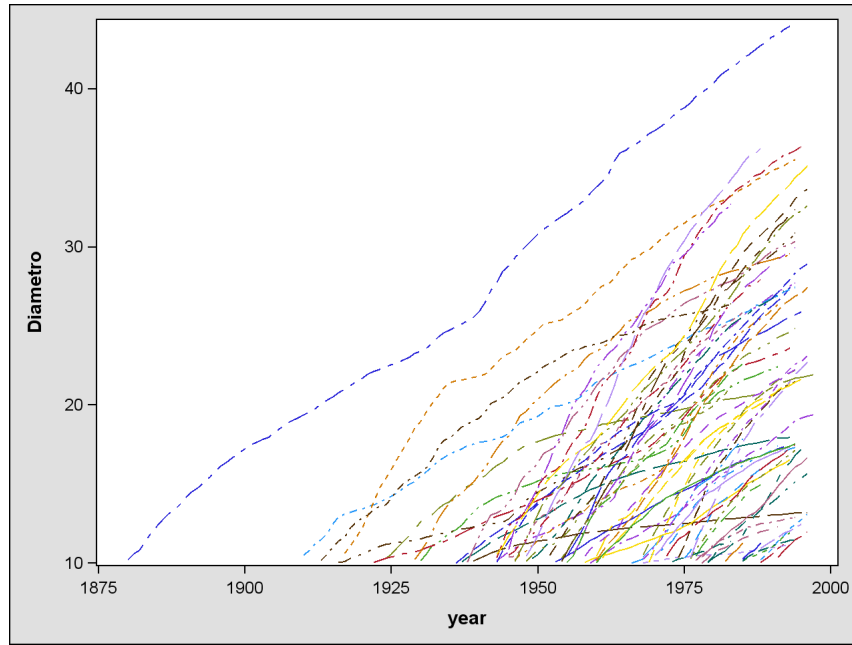
El resumen de los datos es el siguiente:

```
proc means data=a mean std n maxdec=2;  
class y4;  
var Diametro;  
run;
```

Analysis Variable : Diametro				
y4	N Obs	Mean	Std Dev	N
0	771	20.31	6.44	771
1	2352	19.67	6.66	2352
2	1694	17.58	5.97	1694
3	499	12.54	2.26	499

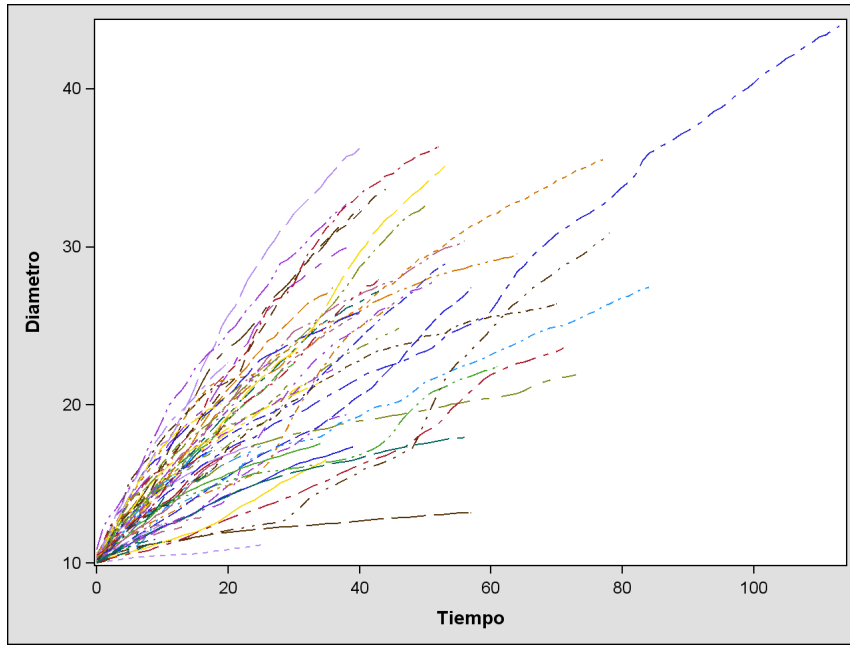
Gráficamente, se presenta una selección de árboles:

```
proc sgplot data=a;  
series y=Diametro x=Year / group=id;  
run;
```



Centrando ahora las observaciones al año de nacimiento:

```
proc sgplot data=a;  
series y=Diametro x=Tiempo / group=id;  
run;
```



El gráfico según período de nacimiento y restringiendo a los primeros 20 años de crecimiento:

```
proc sgpanel data=a;  
panelby y4;  
where tiempo<20;  
series x=Tiempo y=Diametro / group=id ;  
run;
```

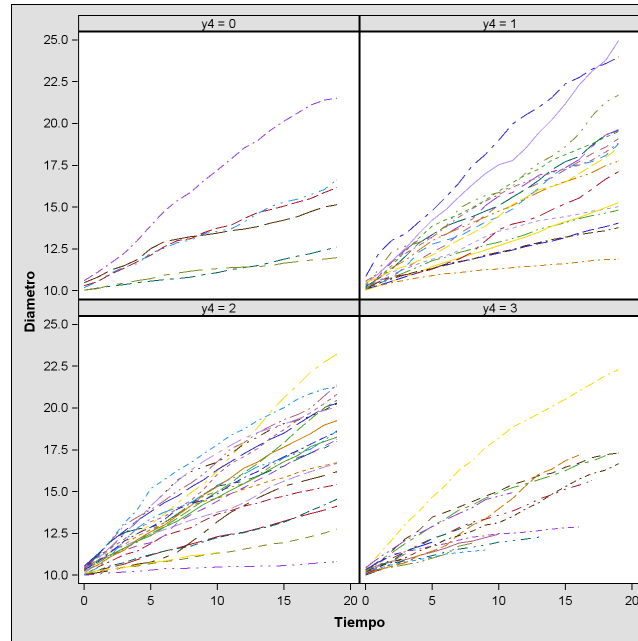
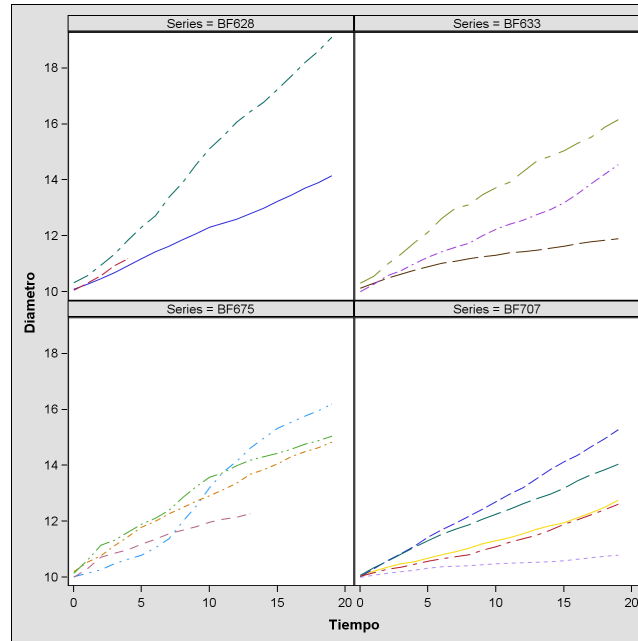


Gráfico representando el crecimiento por parcela (sólo un grupo de parcelas seleccionadas):

```
proc sgpanel data=a;  
panelby parcela;  
where tiempo<20;  
series x=Tiempo y=Diametro / group=id ;  
run;
```



El modelo actual comprende un factor aleatorio, la parcela y anidado, el árbol. Por tratarse de datos longitudinales donde el pendiente de crecimiento varía ligeramente individuo a individuo, es pertinente considerar un modelo mixto con términos independientes aleatorios y pendientes aleatorias. Además, también se considerará un patrón de efectos aleatorios asociado a la parcela donde se ubican los árboles. De esta manera la parcela actúa como factor bloqueado que en cada parcela hay árboles con distinto periodo de nacimiento. El análisis se ve restringido a los primeros 20 años de crecimiento, donde este crecimiento es lineal.

```
proc mixed data=a;
where tiempo<20;
class y4 parcela id;
model Diametro = y4 Tiempo y4*Tiempo / solution dfm=sat residual;
random int Tiempo / subject=parcela ;
random int Tiempo / subject=id(parcela) ;
run;
```

Los índices de ajuste del modelo son:

Fit Statistics	
-2 Res Log Likelihood	3162.6
AIC (smaller is better)	3172.6
AICC (smaller is better)	3172.6
BIC (smaller is better)	3180.3

Las componentes de la varianza y la tabla ANOVA para los efectos fijos son:

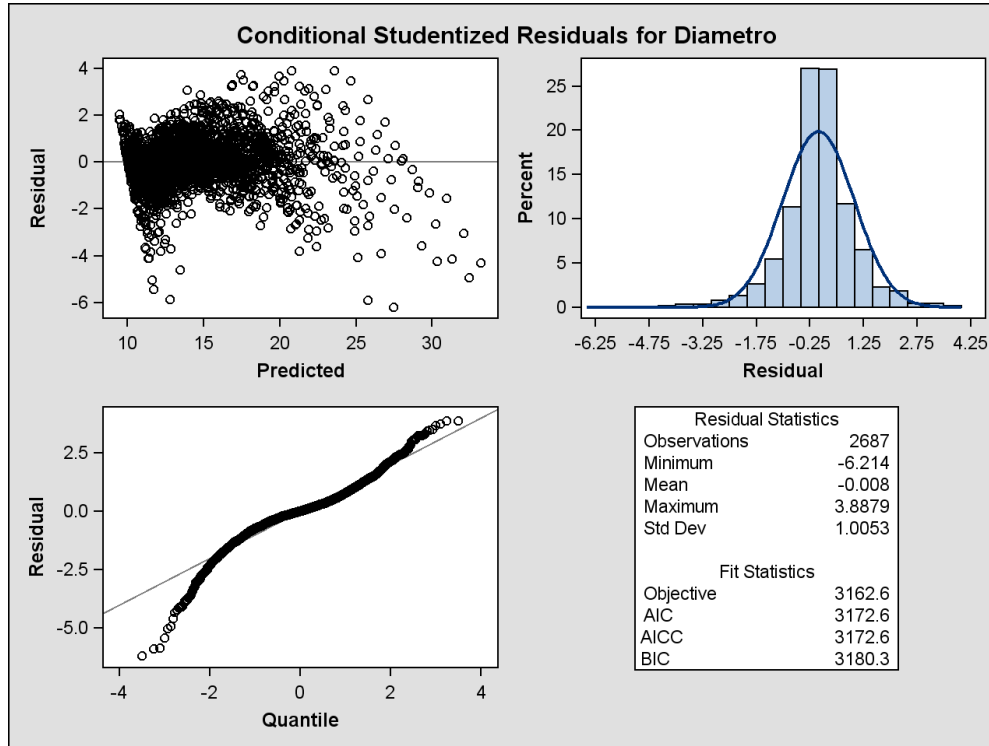
Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	parcela	0.05667
Tiempo	parcela	0.01183
Intercept	id(parcela)	0.2574
Tiempo	id(parcela)	0.03002
Residual		0.1151

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
y4	3	136	1.54	0.2081
Tiempo	1	47.6	213.22	<.0001
Tiempo*y4	3	130	3.83	0.0115

Las estimaciones de los parámetros:

Solution for Fixed Effects						
Effect	y4	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		10.3032	0.09558	126	107.80	<.0001
y4	0	0.09888	0.2023	148	0.49	0.6257
y4	1	0.1977	0.1178	130	1.68	0.0958
y4	2	0.2288	0.1137	134	2.01	0.0461
y4	3	0
Tiempo		0.3109	0.03446	109	9.02	<.0001
Tiempo*y4	0	0.06097	0.06884	142	0.89	0.3773
Tiempo*y4	1	0.08534	0.03931	126	2.17	0.0318
Tiempo*y4	2	0.1275	0.03801	128	3.35	0.0010
Tiempo*y4	3	0

Finalmente, el diagrama de residuos condicionales studentizados es:



Mediante el análisis del diagrama de residuos se observa que podría existir una componente cuadrática no incluida. El modelo se analiza de nuevo añadiendo los nuevos elementos.

El modelo corregido queda como sigue:

```
proc mixed data=a;
where tiempo<20;
class y4 parcela id;
model Diametro = y4 Tiempo y4*Tiempo Tiempo*tiempo y4*tiempo*tiempo / solution dfm=sat residual;
random int Tiempo Tiempo*tiempo / subject=parcela ;
random int Tiempo Tiempo*tiempo / subject=id(parcela) ;
lsmeans y4 / at Tiempo= 10 adjust=tukey;
run;
```

Los índices de ajuste del modelo son:

Fit Statistics	
-2 Res Log Likelihood	201.5
AIC (smaller is better)	215.5
AICC (smaller is better)	215.6
BIC (smaller is better)	226.4

Las componentes de la varianza y la tabla ANOVA para los efectos fijos son:

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	parcela	0.006772
Tiempo	parcela	0.03269
Tiempo*Tiempo	parcela	0.000023
Intercept	id(parcela)	0.09214
Tiempo	id(parcela)	0.06011
Tiempo*Tiempo	id(parcela)	0.000065
Residual		0.02703

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
y4	3	140	1.37	0.2557
Tiempo	1	46.8	121.32	<.0001
Tiempo*y4	3	128	2.96	0.0346
Tiempo*Tiempo	1	51.1	8.66	0.0049
Tiempo*Tiempo*y4	3	127	0.27	0.8437

Las estimaciones de los parámetros:

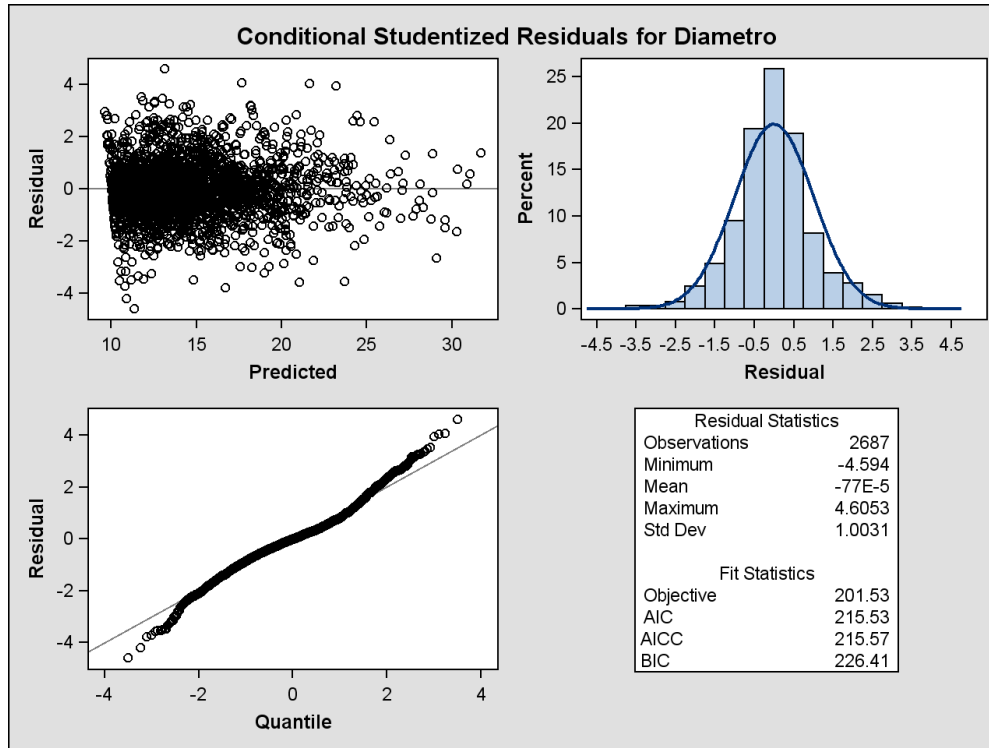
Solution for Fixed Effects						
Effect	y4	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		10.2036	0.05330	135	191.42	<.0001
y4	0	0.01619	0.1171	148	0.14	0.8902
y4	1	0.1373	0.07035	132	1.95	0.0530
y4	2	0.06914	0.06758	138	1.02	0.3081
y4	3	0
Tiempo		0.3580	0.05086	96.7	7.04	<.0001
Tiempo*y4	0	0.08589	0.09882	140	0.87	0.3863
Tiempo*y4	1	0.09184	0.05529	123	1.66	0.0992
Tiempo*y4	2	0.1593	0.05348	126	2.98	0.0035
Tiempo*y4	3	0
Tiempo*Tiempo		-0.00388	0.001833	142	-2.12	0.0359
Tiempo*Tiempo*y4	0	0.000630	0.003337	135	0.19	0.8504
Tiempo*Tiempo*y4	1	0.001081	0.002060	137	0.52	0.6007
Tiempo*Tiempo*y4	2	-0.00043	0.002006	143	-0.21	0.8310
Tiempo*Tiempo*y4	3	0

Finalmente, la tabla de medias ajustadas LSMEANS en el instante de Tiempo = 10 y las comparaciones a posteriori aplicando la corrección de Tukey por multiplicidad de contrastes:

Least Squares Means							
Effect	y4	Tiempo	Estimate	Standard Error	DF	t Value	Pr > t
y4	0	10.00	14.3333	0.9708	174	14.76	<.0001
y4	1	10.00	14.5591	0.5077	103	28.68	<.0001
y4	2	10.00	15.0149	0.4846	88.2	30.99	<.0001
y4	3	10.00	13.3952	0.5301	113	25.27	<.0001

Differences of Least Squares Means										
Effect	y4	_y4	Tiempo	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
y4	0	1	10.00	-0.2257	1.0095	165	-0.22	0.8233	Tukey-Kramer	0.9960
y4	0	2	10.00	-0.6815	1.0064	167	-0.68	0.4992	Tukey-Kramer	0.9056
y4	0	3	10.00	0.9381	1.0387	169	0.90	0.3677	Tukey-Kramer	0.8032
y4	1	2	10.00	-0.4558	0.5417	145	-0.84	0.4015	Tukey-Kramer	0.8346
y4	1	3	10.00	1.1638	0.5806	147	2.00	0.0468	Tukey-Kramer	0.1912
y4	2	3	10.00	1.6196	0.5610	149	2.89	0.0045	Tukey-Kramer	0.0231

Finalmente, el diagrama de residuos condicionales studentizados es:



1.3. Ansiolíticos (2)

En este ejemplo se revisita el ejemplo sobre el análisis de un tratamiento ansiolítico. Ahora, se aplicará la transformación logarítmica a los datos. Al mismo tiempo, dada la naturaleza de las medidas repetidas, otras alternativas para la estructura de covarianzas serán evaluadas.

Ahora, el modelo mixto para analizar el estudio, se basará en un modelo tipo R. De esta manera, se modelizará directamente la estructura de correlaciones. Dicha estructura será común para cada uno de los niveles del factor aleatorio (individuos).

De momento, sólo se explorarán las diferentes propuestas de estructura de correlaciones con el objetivo de validar qué propuesta se ajusta más a los datos obtenidos.

Estructura Compound Symmetry: CS o 1FA

Esta estructura corresponde al mismo tipo de estructura obtenida considerando únicamente el individuo como factor aleatorio.

```
proc mixed data=a;  
class trat vol cond;  
model lss = trat cond trat*cond / dfm=sat residual solution;  
repeated / subject=vol type=cs r=1 rcorr=1;  
run;
```

Los índices de ajuste son los siguientes:

Fit Statistics	
-2 Res Log Likelihood	138.5
AIC (smaller is better)	142.5
AICC (smaller is better)	142.7
BIC (smaller is better)	143.1

La estructura de varianzas resultante es la siguiente:

Estimated R Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.8645	0.6585	0.6585	0.6585	0.6585	0.6585	0.6585	0.6585
2	0.6585	0.8645	0.6585	0.6585	0.6585	0.6585	0.6585	0.6585
3	0.6585	0.6585	0.8645	0.6585	0.6585	0.6585	0.6585	0.6585
4	0.6585	0.6585	0.6585	0.8645	0.6585	0.6585	0.6585	0.6585
5	0.6585	0.6585	0.6585	0.6585	0.8645	0.6585	0.6585	0.6585
6	0.6585	0.6585	0.6585	0.6585	0.6585	0.8645	0.6585	0.6585
7	0.6585	0.6585	0.6585	0.6585	0.6585	0.6585	0.8645	0.6585
8	0.6585	0.6585	0.6585	0.6585	0.6585	0.6585	0.6585	0.8645

y la estructura de correlaciones es:

Estimated R Correlation Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.7617	0.7617	0.7617	0.7617	0.7617	0.7617	0.7617
2	0.7617	1.0000	0.7617	0.7617	0.7617	0.7617	0.7617	0.7617
3	0.7617	0.7617	1.0000	0.7617	0.7617	0.7617	0.7617	0.7617
4	0.7617	0.7617	0.7617	1.0000	0.7617	0.7617	0.7617	0.7617
5	0.7617	0.7617	0.7617	0.7617	1.0000	0.7617	0.7617	0.7617
6	0.7617	0.7617	0.7617	0.7617	0.7617	1.0000	0.7617	0.7617
7	0.7617	0.7617	0.7617	0.7617	0.7617	0.7617	1.0000	0.7617
8	0.7617	0.7617	0.7617	0.7617	0.7617	0.7617	0.7617	1.0000

Estructura UN (sin estructura)

Esta estructura permite considerar un parámetro diferente para cada uno de los elementos de la matriz de varianzas y covarianzas. En este sentido, es una estructura heterogénea. La estructura UN es por lo tanto sumamente flexible, no obstante requiere un elevado número de parámetros.

```
proc mixed data=a;  
class trat vol cond;  
model lss = trat cond trat*cond / dfm=sat residual solution;  
repeated / subject=vol type=un r=1 rcorr=1;  
run;
```

Los índices de ajuste son los siguientes:

Fit Statistics	
-2 Res Log Likelihood	36.0
AIC (smaller is better)	108.0
AICC (smaller is better)	184.2
BIC (smaller is better)	118.9

Las estructuras de varianzas y correlaciones resultantes son las siguientes:

Estimated R Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.5729	0.5348	0.4203	0.2819	0.6487	0.6960	0.6392	0.4692
2	0.5348	0.6864	0.4344	0.3752	0.6743	0.9330	0.7367	0.5926
3	0.4203	0.4344	0.7500	0.5001	0.7052	0.7438	0.8314	0.5790
4	0.2819	0.3752	0.5001	0.4468	0.4451	0.6203	0.6710	0.5795
5	0.6487	0.6743	0.7052	0.4451	1.0049	1.0362	0.8854	0.5226
6	0.6960	0.9330	0.7438	0.6203	1.0362	1.3865	1.1228	0.8425
7	0.6392	0.7367	0.8314	0.6710	0.8854	1.1228	1.1558	0.9155
8	0.4692	0.5926	0.5790	0.5795	0.5226	0.8425	0.9155	0.9125

Estimated R Correlation Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.8529	0.6412	0.5572	0.8549	0.7810	0.7856	0.6489
2	0.8529	1.0000	0.6055	0.6775	0.8120	0.9564	0.8272	0.7488
3	0.6412	0.6055	1.0000	0.8638	0.8122	0.7293	0.8930	0.6998
4	0.5572	0.6775	0.8638	1.0000	0.6642	0.7880	0.9337	0.9075
5	0.8549	0.8120	0.8122	0.6642	1.0000	0.8779	0.8216	0.5457
6	0.7810	0.9564	0.7293	0.7880	0.8779	1.0000	0.8870	0.7490
7	0.7856	0.8272	0.8930	0.9337	0.8216	0.8870	1.0000	0.8914
8	0.6489	0.7488	0.6998	0.9075	0.5457	0.7490	0.8914	1.0000

Estructura con 2 factores aleatorios (2FA)

El diseño del presente estudio contempla el estudio de 4 dosis de tratamiento en sendas visitas. En cada visita se realizan dos mediciones (aunque en condiciones experimentales diferentes). Este diseño contempla en el fondo la presencia de dos factores aleatorios anidados: el individuo y la visita. En cierto sentido se podría considerar que todos los datos procedentes del mismo individuo mostrarán cierto grado de correlación, no obstante, los datos de una misma visita podrían mostrar una asociación mayor que los datos de diferentes visitas. Por este motivo, parece razonable considerar los factores aleatorios VOL y VOL*TRAT. Aunque el ajuste del modelo se realiza empleando un modelo tipo G, existe de forma implícita el modelo tipo R correspondiente.

```
proc mixed data=a;  
class trat vol cond;  
model lss = trat cond trat*cond / dfm=sat residual solution;  
random vol vol*trat / subject=vol v=1 vcorr=1;  
run;
```

Los índices de ajuste son los siguientes:

Fit Statistics	
-2 Res Log Likelihood	130.6
AIC (smaller is better)	136.6
AICC (smaller is better)	136.9
BIC (smaller is better)	137.5

Las estructuras de varianzas y correlaciones resultantes son las siguientes:

Estimated V Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.8645	0.7466	0.6438	0.6438	0.6438	0.6438	0.6438	0.6438
2	0.7466	0.8645	0.6438	0.6438	0.6438	0.6438	0.6438	0.6438
3	0.6438	0.6438	0.8645	0.7466	0.6438	0.6438	0.6438	0.6438
4	0.6438	0.6438	0.7466	0.8645	0.6438	0.6438	0.6438	0.6438
5	0.6438	0.6438	0.6438	0.6438	0.8645	0.7466	0.6438	0.6438
6	0.6438	0.6438	0.6438	0.6438	0.7466	0.8645	0.6438	0.6438
7	0.6438	0.6438	0.6438	0.6438	0.6438	0.6438	0.8645	0.7466
8	0.6438	0.6438	0.6438	0.6438	0.6438	0.6438	0.7466	0.8645

Estimated V Correlation Matrix for Vol 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.8637	0.7447	0.7447	0.7447	0.7447	0.7447	0.7447
2	0.8637	1.0000	0.7447	0.7447	0.7447	0.7447	0.7447	0.7447
3	0.7447	0.7447	1.0000	0.8637	0.7447	0.7447	0.7447	0.7447
4	0.7447	0.7447	0.8637	1.0000	0.7447	0.7447	0.7447	0.7447
5	0.7447	0.7447	0.7447	0.7447	1.0000	0.8637	0.7447	0.7447
6	0.7447	0.7447	0.7447	0.7447	0.8637	1.0000	0.7447	0.7447
7	0.7447	0.7447	0.7447	0.7447	0.7447	0.7447	1.0000	0.8637
8	0.7447	0.7447	0.7447	0.7447	0.7447	0.7447	0.8637	1.0000

Comparación de estructuras

Comparación de estructuras			
Estructura	AIC	AICC	BIC
CS o 1FA	142.5	142.7	143.1
UN	108.0	184.2	118.9
2 FA	136.6	136.9	137.5

El modelo *UN* es la mejor propuesta según los criterios AIC y BIC. Hay que decir también que el modelo seleccionado podría ser también el modelo *2 FA*, el mejor según el criterio AICC y con sentido práctico ya que contempla el diseño del experimento.

Ajuste del modelo final: estructura UN

El modelo contemplando una estructura de correlaciones del tipo UN se vuelve a analizar.

```
proc mixed data=a;
class trat vol cond;
model lss = trat cond trat*cond / dfm=sat residual ;
repeated / subject=vol type=un r=1 rcorr=1;
run;
```

Test LRT de bondad de ajuste:

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
35	176.22	<.0001

Tabla ANOVA para los efectos fijos y estimaciones de los parámetros:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Trat	3	9	18.29	0.0004
Cond	1	9	19.08	0.0018
Trat*Cond	3	9	1.84	0.2099

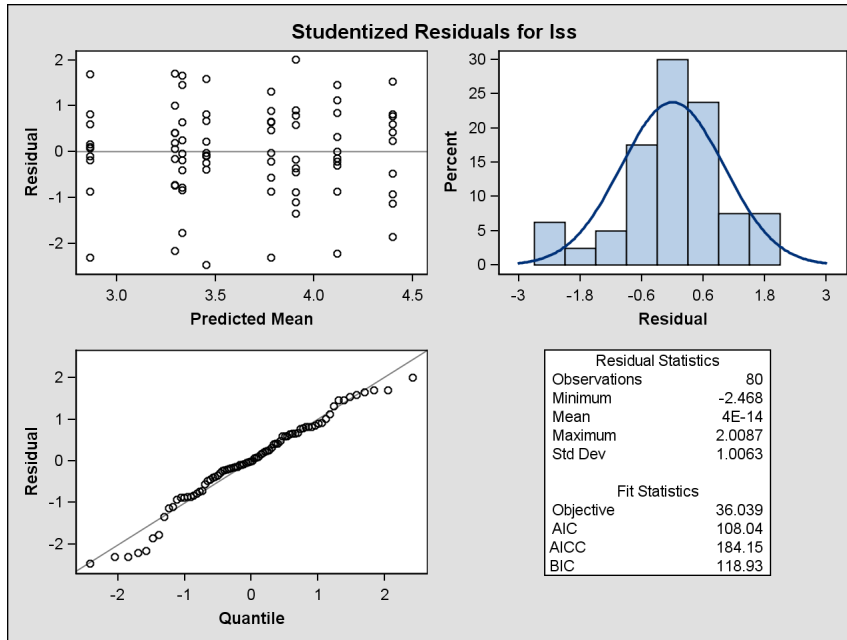
Solution for Fixed Effects							
Effect	Trat	Cond	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			3.3308	0.3021	9	11.03	<.0001
Trat	1		1.0680	0.2034	9	5.25	0.0005
Trat	2		0.7875	0.1415	9	5.56	0.0003
Trat	3		0.4530	0.2478	9	1.83	0.1008
Trat	4		0
Cond		1	-0.4667	0.1541	9	-3.03	0.0143
Cond		2	0
Trat*Cond	1	1	-0.02426	0.1937	9	-0.13	0.9031
Trat*Cond	1	2	0
Trat*Cond	2	1	-0.1980	0.1058	9	-1.87	0.0942
Trat*Cond	2	2	0
Trat*Cond	3	1	-0.02198	0.1978	9	-0.11	0.9139
Trat*Cond	3	2	0
Trat*Cond	4	1	0
Trat*Cond	4	2	0

Tabla de medias ajustadas LSMEANS y las comparaciones a posteriori aplicando la corrección de Tukey por multiplicidad de contrastes:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Trat	3	9	18.29	0.0004
Cond	1	9	19.08	0.0018
Trat*Cond	3	9	1.84	0.2099

Least Squares Means						
Effect	Trat	Estimate	Standard Error	DF	t Value	Pr > t
Trat	1	4.1533	0.2413	9	17.21	<.0001
Trat	2	3.7859	0.2344	9	16.15	<.0001
Trat	3	3.5394	0.3341	9	10.60	<.0001
Trat	4	3.0974	0.3122	9	9.92	<.0001

Con el objetivo de validar las suposiciones de Normalidad de los residuos, podemos obtener el diagrama de residuos marginales estudentizados:



No es posible obtener los residuos condicionales puesto que se trata de un modelo tipo R

1.4. Episodios Enfermedad

En este caso, se desea analizar el patrón de evolución de una patología auto-inmune. La información de los pacientes es recogida semestralmente. La variable objetivo consiste en el número de brotes experimentados desde la última visita realizada. Se selecciona únicamente a aquellos pacientes con al menos 3 visitas realizadas, en total se incluyen 277 sujetos. Finalmente, se desea analizar la influencia de la edad en el momento de diagnóstico o del sexo en la severidad de la enfermedad.

A continuación se muestran algunos datos:

```
proc print data=a(obs=10);
run;
```

Obs	CodiPacient	Sexo	Edad	NBrotos	Tiempo	nbrotos	n	Tiempo2	Edad2	EdadR	visita	id
1	1503	1	53	0	238	0	9	0.06516	0.13	50	1	1
2	1503	1	53	0	420	0	9	0.11499	0.13	50	2	1
3	1503	1	53	0	728	0	9	0.19932	0.13	50	3	1
4	1503	1	53	0	791	0	9	0.21656	0.13	50	4	1
5	1503	1	53	0	961	0	9	0.26311	0.13	50	5	1
6	1503	1	53	0	1156	0	9	0.31650	0.13	50	6	1
7	1503	1	53	0	1463	0	9	0.40055	0.13	50	7	1
8	1503	1	53	0	1610	0	9	0.44079	0.13	50	8	1
9	1504	0	53	0	279	0	16	0.07639	0.13	50	1	2
10	1504	0	53	0	448	0	16	0.12266	0.13	50	2	2

El resumen de los datos es el siguiente:

```
proc means data=a mean std n maxdec=2;
class sexo;
var nbrotos;
run;
```

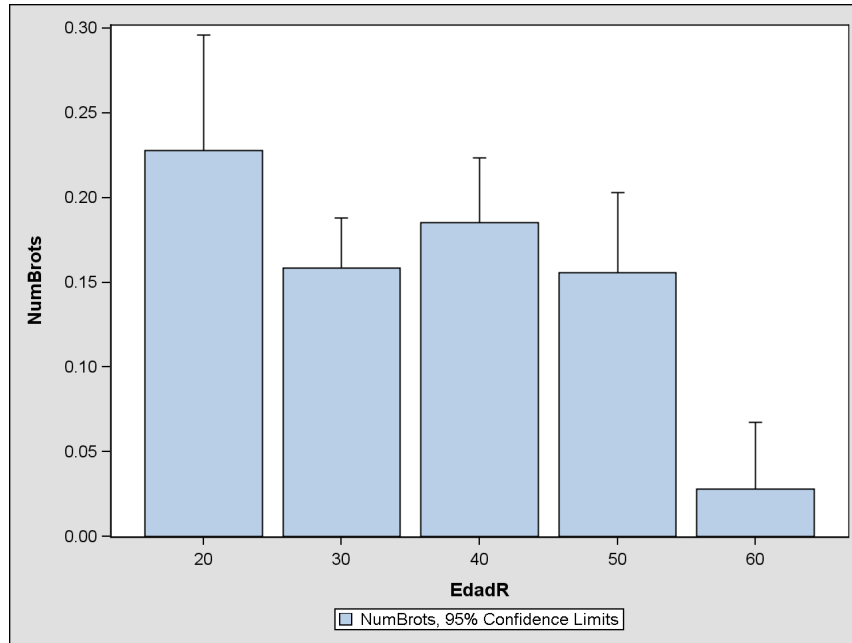
Analysis Variable : NBrotos NumBrots				
Sexe	N Obs	Mean	Std Dev	N
0	518	0.15	0.38	518
1	1301	0.18	0.45	1301

```
proc means data=a mean std n maxdec=2;
class EdadR;
var nbrotos;
run;
```

Analysis Variable : NBrotos NumBrots				
EdadR	N Obs	Mean	Std Dev	N
20	272	0.23	0.57	272
30	682	0.16	0.40	682
40	486	0.19	0.43	486
50	308	0.16	0.42	308
60	71	0.03	0.17	71

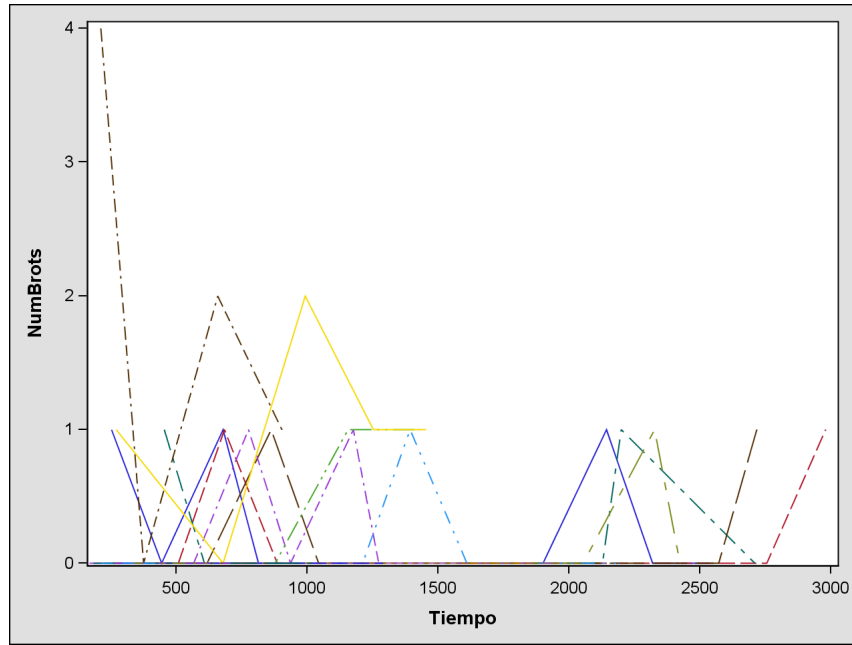
Gráficamente:

```
proc sgplot data=a;  
vbar edad / response=nbrotos stat=mean limits=upper ;  
run;
```



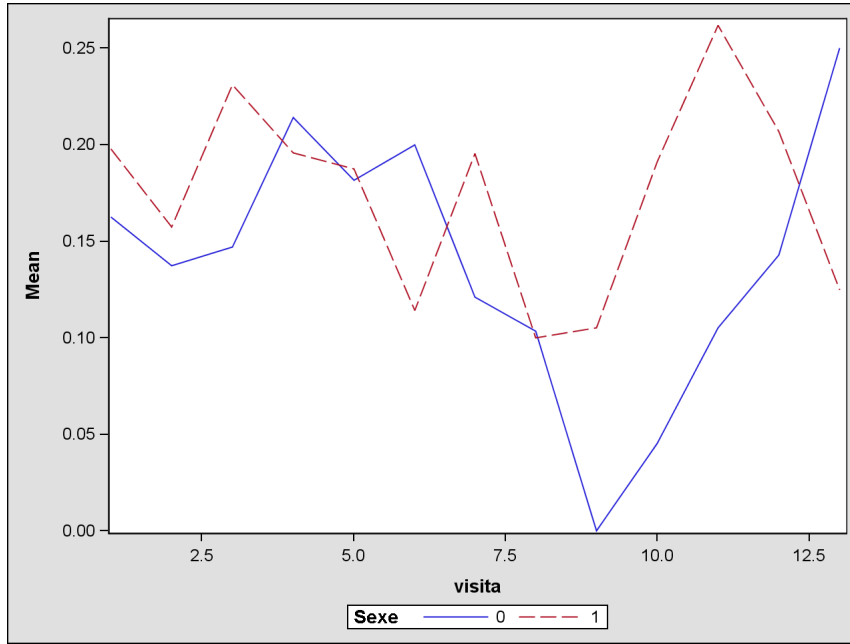
La evolución a lo largo del tiempo a nivel individual para una selección de sujetos es la siguiente:

```
proc sgplot data=a noautolegend;  
series x=tiempo y=nbrotos / group=id ;  
run;
```



La evolución promedio a nivel visita según el sexo se muestra en el siguiente gráfico:

```
proc sgplot data=s;  
series x=visita y=nbrotesm / group=sexo ;  
run;
```



El modelo actual comprende un factor aleatorio, el individuo. Dado que los datos son recuentos, se empleará un modelo lineal generalizado mixto. Por tratarse de datos longitudinales donde la evolución de la patología puede variar, se considerará el término independiente y el pendiente asociados a cada individuo como efectos aleatorios.

```
proc glimmix data=a plots=pearsonpanel maxopt=50 method=quad;
class sexo id;
model nbrotos = edadr sexo Tiempo / dist=poisson ;
random intercept Tiempo / subject=id;
run;
```

Fit Statistics	
-2 Log Likelihood	1698.54
AIC (smaller is better)	1710.54
AICC (smaller is better)	1710.59
BIC (smaller is better)	1732.29
CAIC (smaller is better)	1738.29
HQIC (smaller is better)	1719.27

Fit Statistics for Conditional Distribution	
-2 log L(NBrotos r. effects)	1431.59
Pearson Chi-Square	1224.52
Pearson Chi-Square / DF	0.67

Las componentes de la varianza y la tabla ANOVA para los efectos fijos son:

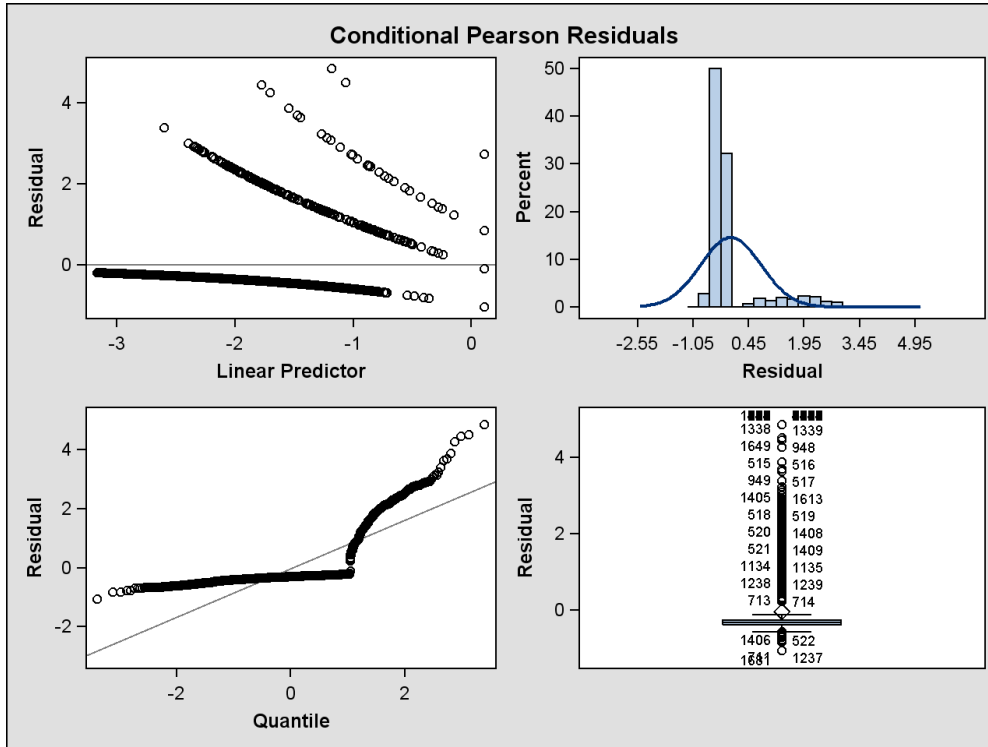
Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	id	0.7582	0.1842
Tiempo2	id	0.4739	1.0680

Solutions for Fixed Effects						
Effect	Sexe	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-1.5056	0.3316	274	-4.54	<.0001
EdadR		-0.01570	0.008560	1264	-1.83	0.0669
Sexo	0	-0.1584	0.1953	1264	-0.81	0.4173
Sexo	1	0
Tiempo2		-0.06938	0.4296	277	-0.16	0.8718

Las estimaciones de los parámetros:

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
EdadR	1	1264	3.36	0.0669
Sexo	1	1264	0.66	0.4173
Tiempo2	1	277	0.03	0.8718

Finalmente, el diagrama de residuos condicionales de Pearson es:



1.5. Deporte de lucha

En el presente estudio, perteneciente al ámbito del Taekwondo, se desea medir y analizar los factores que influyen en la cantidad de movimientos ofensivos realizados por los contrincantes así como su éxito. En este deporte se enfrentan dos participantes que realizan sucesivamente movimientos ofensivos o defensivos. Cada combate se divide en 3 asaltos de 3 minutos cada uno. Un juez-árbitro puntúa los movimientos ofensivos y otorga la victoria al jugador que alcanza un mayor número de puntos. Se dispone de datos correspondientes a 32 combates diferentes pertenecientes a distintas categorías de peso (Pluma, Mini-Mosca, etc.) obtenidos en 5 competiciones oficiales distintas. En cada combate, se recogieron datos de los dos participantes. No obstante, algunos participantes han sido evaluados en más de un combate, en este sentido, hay 49 deportistas diferentes. En cada uno de los combates se ha recogido el éxito o fracaso de cada movimiento ofensivo realizado, así como diferentes características sobre la técnica empleada, el tiempo transcurrido de cada asalto o el estado del marcador. En total, se dispone de 2246 movimientos ofensivos. Los datos han sido agrupados según intervalos de tiempo de 1 minuto de duración obteniendo dos variables objetivo: el recuento de movimientos ofensivos realizados por minuto y el recuento de movimientos con éxito por minuto. A continuación se muestran algunos datos:

```
proc print data=a(obs=10);
run;
```

Obs	Peso	combate	id	asalto	t	intentos	exitos	Marcador	Desequilibrio
1	Gallo	1	1	1	0	2	1	0	0
2	Gallo	1	2	1	0	2	0	0	0
3	Gallo	1	1	1	1	3	0	1	1
4	Gallo	1	2	1	1	3	0	-1	1
5	Gallo	1	1	1	2	2	1	1	1
6	Gallo	1	2	1	2	4	1	-1	1
7	Gallo	1	1	2	0	2	0	1	1
8	Gallo	1	2	2	0	3	1	-1	1
9	Gallo	1	1	2	1	3	0	0	0
10	Gallo	1	2	2	1	2	0	0	0
11	Gallo	1	1	2	2	1	0	0	0
12	Gallo	1	2	2	2	3	0	0	0
13	Gallo	1	1	3	0	2	0	0	0
14	Gallo	1	2	3	0	3	0	0	0
15	Gallo	1	1	3	1	8	1	0	0
16	Gallo	1	2	3	1	4	0	0	0
17	Gallo	1	1	3	2	7	1	1	1
18	Gallo	1	2	3	2	7	2	-1	1

El resumen de los datos es el siguiente:

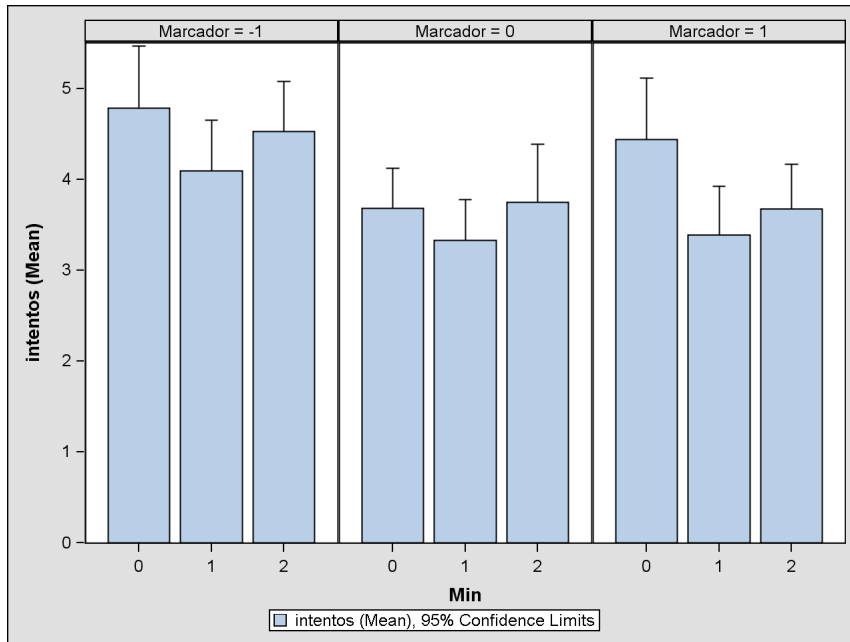
```
proc means data=a mean std n maxdec=2;  
class marcador ;  
var exitos intentos;  
run;
```

Analysis Variable : intentos				
Asalto	N Obs	Mean	Std Dev	N
1	192	3.55	2.22	192
2	192	3.76	2.05	192
3	192	4.39	2.30	192

Analysis Variable : exitos				
Asalto	N Obs	Mean	Std Dev	N
1	192	0.61	0.78	192
2	192	0.67	0.88	192
3	192	0.80	0.94	192

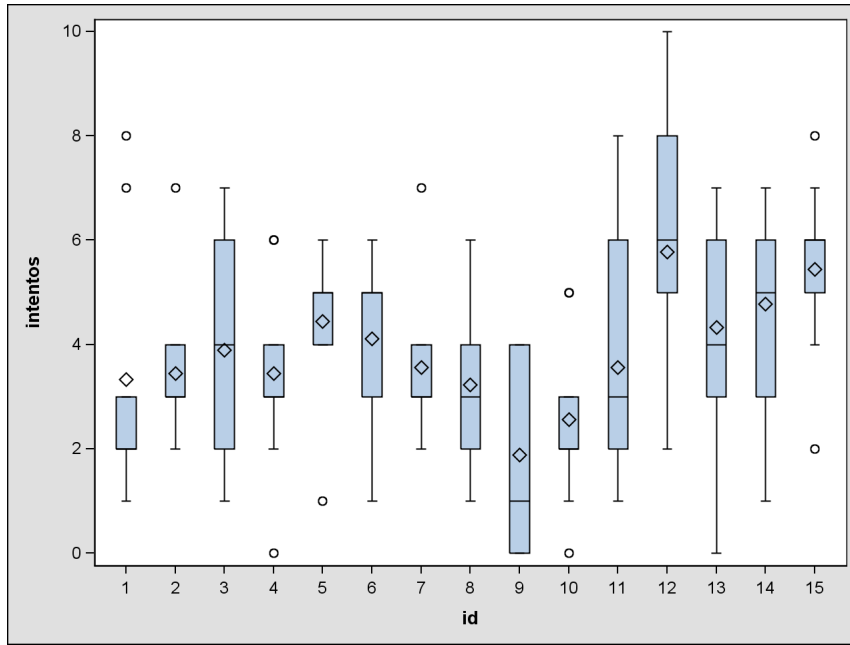
Gráficamente:

```
proc sgpanel data=a;  
panelby marcador;  
vbar t / response=intentos stat=mean limits=upper ;  
run;
```



Puede existir cierta heterogeneidad entre participantes, gráficamente:

```
proc sgplot data=a noautolegend;  
vbox id / response=intentos stat=mean limits=upper group=peso ;  
run;
```



El modelo actual comprende dos factores aleatorios, el combate y el individuo. En cada combate se realizan una serie de mediciones, al mismo tiempo cada sujeto es evaluado en un número de movimientos. Existe cierta heterogeneidad entre individuos y entre combates. No obstante, el factor combate será despreciado del modelo debido a su poca contribución. Se considerará cada individuo a partir de la combinación como sujeto - combate.

Inicialmente se analizará el recuento de movimientos ofensivos independientemente de su éxito. En este sentido, se empleará un modelo lineal generalizado mixto para variables respuesta poisson contemplando un sólo factor aleatorio, el individuo.

```
proc glimmix data=a pconv=1e-5 plots=pearsonpanel;
class combate id marcador t asalto peso;
model intentos = peso desequilibrio marcador t asalto / dist=poisson solution;
random intercept / subject=id*combate v=1 vcorr=1 ;
lsmeans asalto / ilink;
run;
```

Criterios de ajuste:

Fit Statistics	
-2 Res Log Pseudo-Likelihood	937.50
Generalized Chi-Square	507.65
Gener. Chi-Square / DF	0.90

Componentes de la varianza:

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	combate*id	0.07766	0.02134

Estructura de varianzas y estructura de correlaciones:

Estimated V Matrix for combate*id 1 1									
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	0.4039	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766
2	0.07766	0.4832	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766
3	0.07766	0.07766	0.4377	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766
4	0.07766	0.07766	0.07766	0.4049	0.07766	0.07766	0.07766	0.07766	0.07766
5	0.07766	0.07766	0.07766	0.07766	0.4031	0.07766	0.07766	0.07766	0.07766
6	0.07766	0.07766	0.07766	0.07766	0.07766	0.3665	0.07766	0.07766	0.07766
7	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.3158	0.07766	0.07766
8	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.3424	0.07766
9	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.07766	0.3405

Estimated V Correlation Matrix for combate*id 1 1									
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.0000	0.1758	0.1847	0.1920	0.1925	0.2018	0.2175	0.2088	0.2094
2	0.1758	1.0000	0.1688	0.1756	0.1760	0.1845	0.1988	0.1909	0.1914
3	0.1847	0.1688	1.0000	0.1845	0.1849	0.1939	0.2089	0.2006	0.2012
4	0.1920	0.1756	0.1845	1.0000	0.1922	0.2016	0.2172	0.2086	0.2092
5	0.1925	0.1760	0.1849	0.1922	1.0000	0.2020	0.2177	0.2090	0.2096
6	0.2018	0.1845	0.1939	0.2016	0.2020	1.0000	0.2283	0.2192	0.2198
7	0.2175	0.1988	0.2089	0.2172	0.2177	0.2283	1.0000	0.2362	0.2368
8	0.2088	0.1909	0.2006	0.2086	0.2090	0.2192	0.2362	1.0000	0.2274
9	0.2094	0.1914	0.2012	0.2092	0.2096	0.2198	0.2368	0.2274	1.0000

Y la tabla ANOVA para los efectos fijos es la siguiente:

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Peso	7	505	1.89	0.0688
Desequilibrio	1	505	9.92	0.0017
Marcador	2	505	6.66	0.0014
t	2	505	3.03	0.0493
asalto	2	505	12.96	<.0001

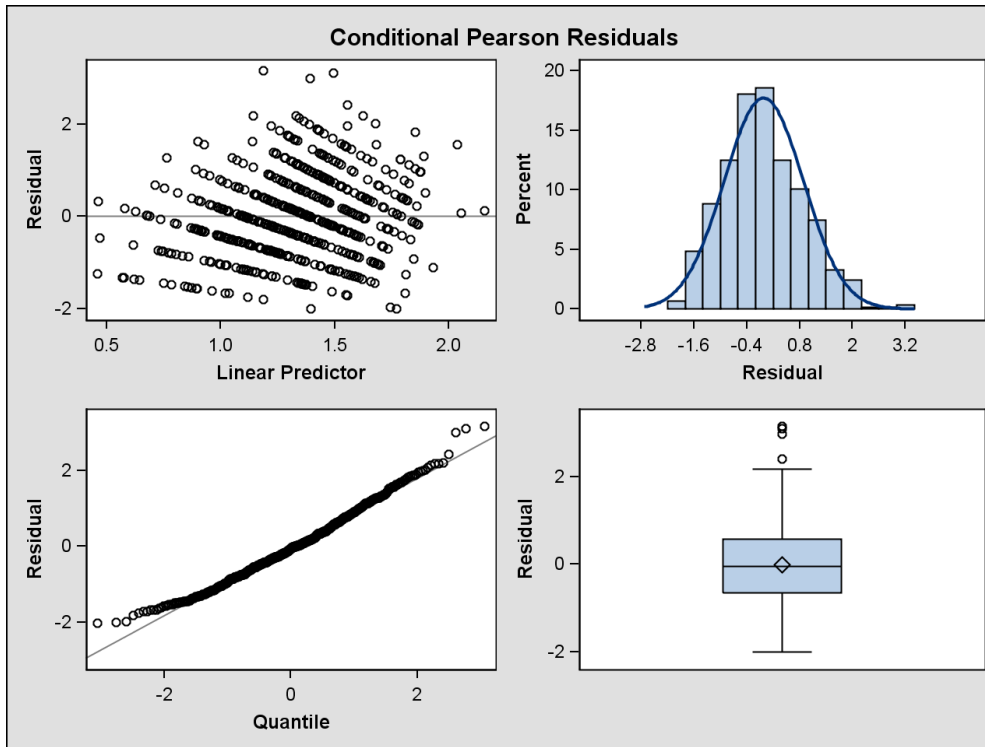
Las estimaciones LSMEANS para la variable Asalto:

asalto Least Squares Means							
Asalto	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
1	1.1957	0.05624	505	21.26	<.0001	3.3059	0.1859
2	1.3043	0.05141	505	25.37	<.0001	3.6849	0.1895
3	1.5105	0.05274	505	28.64	<.0001	4.5291	0.2389

Las estimaciones de los parámetros:

Solutions for Fixed Effects									
Effect	Peso	Marcador	Min	Asalto	Estimate	Standard Error	DF	t Value	Pr > t
Intercept					1.7425	0.1581	56	11.02	<.0001
Peso	Gallo				-0.3402	0.1695	505	-2.01	0.0453
Peso	Ligero				-0.1296	0.1744	505	-0.74	0.4580
Peso	Medio				-0.08508	0.1740	505	-0.49	0.6251
Peso	MiniMosca				0.1267	0.1801	505	0.70	0.4823
Peso	Mosca				-0.3535	0.1772	505	-1.99	0.0466
Peso	Pesado				-0.3001	0.1768	505	-1.70	0.0903
Peso	Pluma				-0.1407	0.1748	505	-0.80	0.4214
Peso	Superligero				0
Desequilibrio					-0.09082	0.02884	505	-3.15	0.0017
Marcador		-1			0.2284	0.06816	505	3.35	0.0009
Marcador		0			0.02084	0.07811	505	0.27	0.7897
Marcador		1			0
t			0		-0.01293	0.05359	505	-0.24	0.8095
t			1		-0.1191	0.05262	505	-2.26	0.0240
t			2		0
asalto				1	-0.3148	0.06443	505	-4.89	<.0001
asalto				2	-0.2063	0.05374	505	-3.84	0.0001
asalto				3	0

Finalmente, el diagrama de residuos condicionales de Pearson es:



Las distintas medidas repetidas realizadas en cada sujeto están estructuradas de forma compacta: todos los sujetos son evaluados en 9 ocasiones por combate, 3 mediciones por asalto. Así, de forma alternativa es posible considerar un modelo lineal mixto generalizado con variable respuesta de Poisson y con una estructura de varianzas particular. Esta visión empleará el método GEE para su análisis y conllevará una interpretación marginal de los resultados. La estructura propuesta será una estructura UN, aunque existen un gran número de propuestas disponibles: CS, CSH, AR(1), ARH(1), etc.

```
proc glimmix data=a empirical plots=pearsonpanel;
class combate id marcador t asalto peso;
model intentos = peso desequilibrio marcador t asalto / dist=poi solution ;
random _residual_ / subject= id*combate type=un vcorr=1 ;
lsmeans asalto / ilink;
run;
```

Criterios de ajuste:

Fit Statistics	
-2 Res Log Pseudo-Likelihood	885.30
Generalized Chi-Square	561.01
Gener. Chi-Square / DF	1.00

Estructura de correlaciones:

Estimated V Correlation Matrix for combate*id 1 1									
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.0000	0.2135	-0.01757	0.4258	0.5185	0.4252	0.4682	0.3763	0.1695
2	0.2135	1.0000	0.01560	0.4822	0.2457	0.1304	0.2964	0.1443	0.05684
3	-0.01757	0.01560	1.0000	0.2349	0.07341	0.02930	0.1514	0.01426	0.2007
4	0.4258	0.4822	0.2349	1.0000	0.3406	0.2896	0.5594	0.09621	0.2937
5	0.5185	0.2457	0.07341	0.3406	1.0000	0.2816	0.4397	0.2845	0.2888
6	0.4252	0.1304	0.02930	0.2896	0.2816	1.0000	0.2346	0.1129	-0.04440
7	0.4682	0.2964	0.1514	0.5594	0.4397	0.2346	1.0000	0.2820	0.3254
8	0.3763	0.1443	0.01426	0.09621	0.2845	0.1129	0.2820	1.0000	0.3627
9	0.1695	0.05684	0.2007	0.2937	0.2888	-0.04440	0.3254	0.3627	1.0000

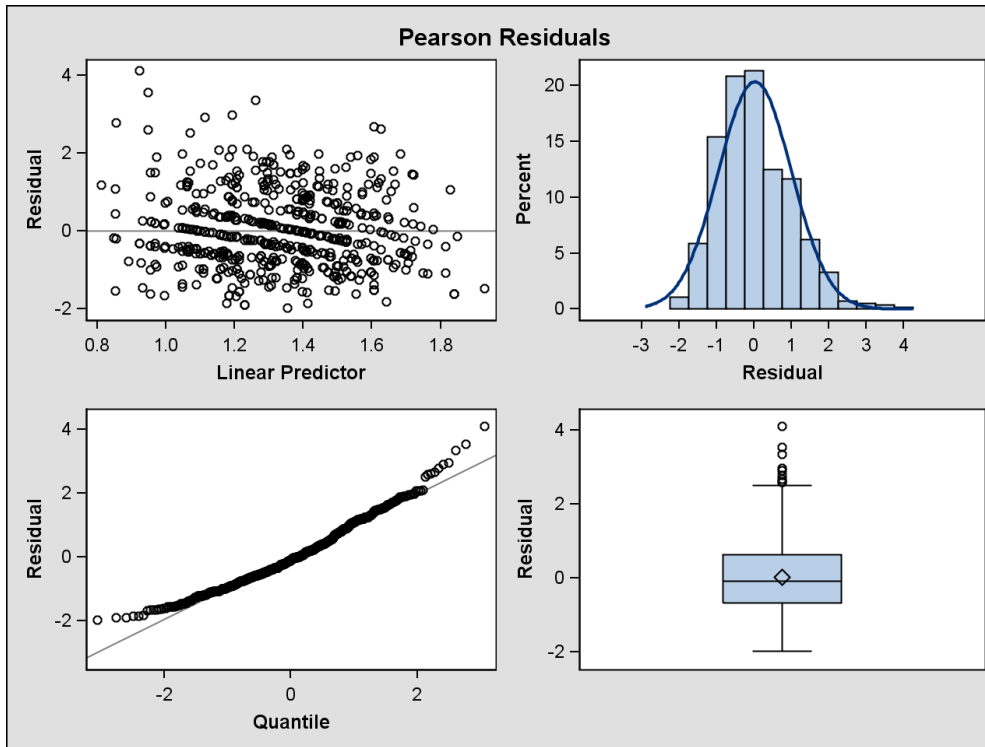
Y la tabla ANOVA para los efectos fijos es la siguiente:

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Peso	7	56	5.96	<.0001
Desequilibrio	1	56	13.53	0.0005
Marcador	2	56	8.67	0.0005
t	2	56	3.43	0.0395
asalto	2	56	12.77	<.0001

Las estimaciones LSMEANS para la variable Asalto:

asalto Least Squares Means							
Asalto	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
1	1.1995	0.05293	56	22.66	<.0001	3.3185	0.1757
2	1.3195	0.04956	56	26.62	<.0001	3.7415	0.1854
3	1.5134	0.04492	56	33.69	<.0001	4.5423	0.2040

Finalmente, el diagrama de residuos marginales de Pearson:



Con el objetivo de analizar la tasa de éxitos, se propone un modelo lineal generalizado mixto con variable respuesta binomial (éxitos / intentos) con efectos aleatorios, aunque también podría considerarse la implementación de cierta estructura de varianzas como en el caso anterior:

```
proc glimmix data=a plots=pearsonpanel;  
class combate id marcador t asalto peso;  
model exitos / intentos = peso desequilibrio marcador t asalto / dist=binomial solution;  
random intercept / subject=id*combate;  
lsmeans peso / ilink;  
run;
```

Criterios de ajuste:

Fit Statistics	
-2 Res Log Pseudo-Likelihood	2085.29
Generalized Chi-Square	568.42
Gener. Chi-Square / DF	1.05

Componentes de la varianza:

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	combate*id	0.2022	0.09355

La tabla ANOVA para los efectos fijos es la siguiente:

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Peso	7	487	2.62	0.0114
Desequilibrio	1	487	1.86	0.1737
Marcador	2	487	0.10	0.9030
t	2	487	0.44	0.6458
asalto	2	487	1.16	0.3134

Las estimaciones LSMEANS para la variable Peso:

Peso Least Squares Means							
Peso	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
Gallo	-1.5285	0.2129	487	-7.18	<.0001	0.1782	0.03118
Ligero	-2.1459	0.2508	487	-8.56	<.0001	0.1047	0.02351
Medio	-1.0094	0.2083	487	-4.85	<.0001	0.2671	0.04078
MiniMosca	-1.9013	0.2762	487	-6.88	<.0001	0.1300	0.03122
Mosca	-1.1779	0.2280	487	-5.17	<.0001	0.2354	0.04104
Pesado	-1.5673	0.2361	487	-6.64	<.0001	0.1726	0.03372
Pluma	-1.5678	0.2256	487	-6.95	<.0001	0.1725	0.03221
Superligero	-1.8683	0.2658	487	-7.03	<.0001	0.1337	0.03079

Una última alternativa para analizar la tasa de éxitos consiste en emplear como variable respuesta la variable EXITOS modelizada como variable Poisson y añadir el logaritmo del número de intentos como término OFFSET.

```
proc glimmix data=a plots=pearsonpanel;  
class combate id marcador t asalto peso;  
model exitos = peso desequilibrio marcador t asalto / dist=poi solution offset=lintentos;  
random intercept / subject=id*combate ;  
run;
```

Criterios de ajuste:

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1865.96
Generalized Chi-Square	482.38
Gener. Chi-Square / DF	0.89

Componentes de la varianza:

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	combate*id	0.07588	0.05656

La tabla ANOVA para los efectos fijos es la siguiente:

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Peso	7	487	2.95	0.0048
Desequilibrio	1	487	1.47	0.2252
Marcador	2	487	0.12	0.8904
t	2	487	0.36	0.6960
asalto	2	487	0.92	0.3980

Finalmente, el diagrama de residuos condicionales de Pearson es:

