Instituto Valenciano de Investigaciones Agrarias

Seminario

MÉTODOS ESTADÍSTICOS PARA LA INVESTIGACIÓN AGRONÓMICA

Tema 7

ANÁLISIS DE LA VARIANZA MULTIFACTORIALES

Análisis de la Varianza Multifactoriales

- Planes Factoriales Equilibrados
- Análisis de la Varianza
- Estudio de los efectos simples
- Estudio de las interacciones dobles
  - Gráficos
  - Descomposición mediante contrastes ortogonales
- Estudio de efectos sobre la varianza
- Ejemplo 1: Efecto de la variedad y de la altura de corte sobre rendimiento de sorgo forrajero (PFE 3x4 con 3 réplicas)
- Ejemplo 2: Estudio de 4 factores para mejorar un proceso de adhesivado (Plan 2^4)
- Ejemplo 3: Efecto del sexo y la temperatura en la longevidad de adultos de *aganaspis daci*

PLANES FACTORIALES EQUILIBRADOS

- Sea un estudio en el que se van a investigar los efectos de K factores sobre una variable respuesta.
  - Factor 1: se plantean n_1 niveles o variantes
  - Factor 2: se plantean n_2 niveles o variantes
  - ……………………………………………
  - Factor K se plantean n_k niveles o variantes
- El número de posibles condiciones diferentes que podrían plantearse será = $n_1 \times n_2 \times \ldots \times n_k$. A cada una de estas condiciones, a las que se denomina tratamiento, le corresponde una determinada población.
- Un Plan Factorial se denomina Equilibrado (o Balanceado) si para cada condición posible se dispone de un mismo número n_r de datos
- Si $n_r = 1$ el Plan Factorial se denomina no replicado

ANÁLISIS DE LA VARIANZA DE PFE

- En el cuadro resumen del Anova el efecto simple de un factor con n_k variantes (o niveles) tiene n_k-1 grados de libertad
- Los grados de libertad de cualquier interacción son el producto de los grados de libertad de los factores que interaccionan
- La SC y los gl residuales se calculan como la diferencia entre los correspondientes totales (los gl_{total} son siempre el número de datos - 1) y las sumas de los de los efectos considerados
- En los planes factoriales **no replicados,** si se consideran todas las interacciones posibles **no puede calcularse directamente el $CM_{residual}$** porque los grados de libertad residuales serían 0.
- En general no se estudian las interacciones de orden mayor que 2, pues rara vez son significativas y resultan difíciles de interpretar. Las SC y los gl asociados a todas estas interacciones se agrupan para obtener una estimación del $CM_{residual}$
- **Si los gl_{resid} son reducidos** (por ejemplo, menores que 10) **los análisis son poco potentes**. Para obviar este problema, se pueden acumular también en el residuo las SC y los g.l. de aquellos efectos que resulten pequeños (por ejemplo, con F_{ratio} menores que 2), especialmente si corresponden a interacciones.

ANOVA si el Plan Factorial no es equilibrado

- En ocasiones, por ejemplo, por pérdida de alguna unidad experimental, el diseño resultante no está perfectamente equilibrado
- El análisis de estos diseños desequilibrados es muy complejo computacionalmente, pero afortunadamente eso no es problema actualmente para el investigador que dispone del software adecuado (Elegir Suma de Cuadrados de Tipo III)
- Si un diseño no está perfectamente equilibrado las estimaciones de los diferentes ejemplos no son totalmente independientes, sino que están parcialmente confundidas entre sí
- Si faltan la totalidad de los datos correspondientes a una determinada combinación de niveles de dos factores es imposible estudiar la interacción doble entre ellos

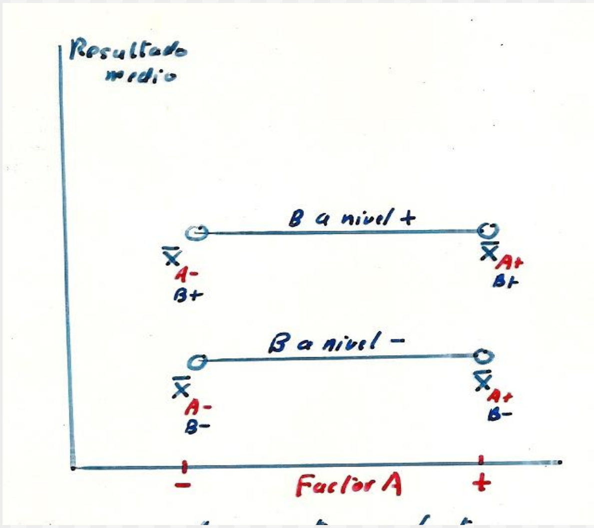
Análisis de los efectos simples

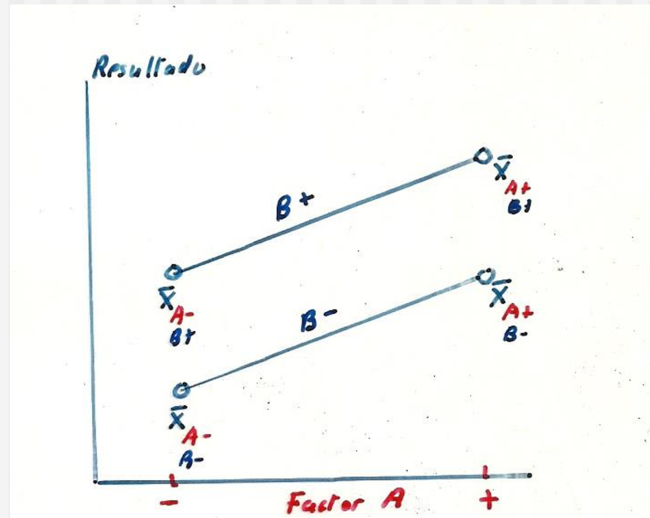
- El análisis de los efectos simples que resulten significativos se lleva a cabo de la forma expuesta en el tema anterior
- Si el factor es de naturaleza cualitativa, la significación de las diferencias entre las medias de las diferentes variantes puede llevarse a cabo mediante intervalos LSD u otros procedimientos de comparaciones múltiples (Tukey, Newman-Keuls, …)
- Si las diferentes variantes del factor cualitativo corresponden a cierta estructura, también puede utilizarse la técnica de contrastes ortogonales
- Si el factor es de naturaleza cuantitativa **siempre** hay que analizarlo estudiando, mediante contrastes ortogonales, la significación de las diferentes componentes lineal, cuadrática, …
- Cuando el efecto de un factor cuantitativo resulte "casi" significativo, conviene estudiar si al menos la componente lineal resulta finalmente significativa

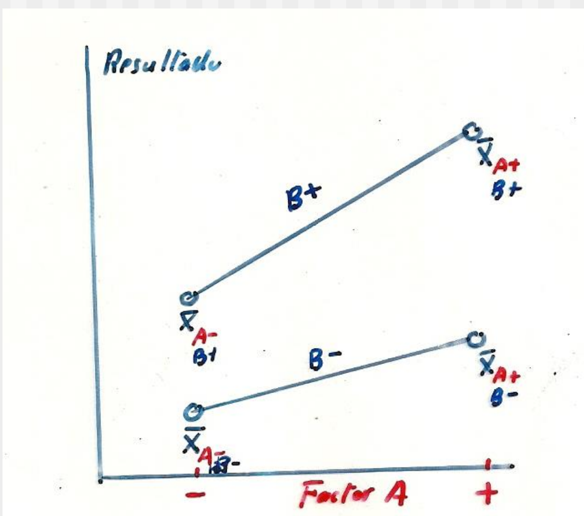
Análisis de las interacciones dobles

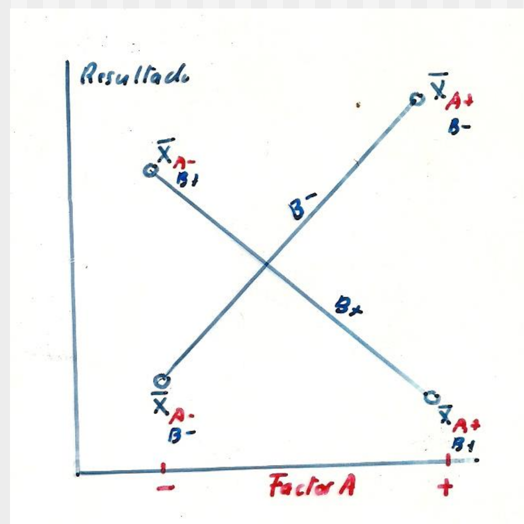
- Toda interacción significativa entre dos factores A y B puede analizarse de dos formas:
  - Estudiando cómo el efecto del factor B es diferente según la variante considerada del factor A
  - Estudiando cómo el efecto del factor A es diferente según la variante considerada del factor B
- Ambas interpretaciones son completamente equivalentes ("dos caras de la misma moneda") pero en ocasiones una de ambas puede resultar más intuitiva o relevante para el objeto de la investigación
- Para interpretar una interacción doble resultan especialmente útiles las **representaciones gráficas**
- En estas representaciones se sitúan en el eje de abscisas las diferentes alternativas para un factor A, y sobre cada una de ellas se grafican los valores medios correspondientes a las diferentes variantes del otro factor B (uniéndose luego con líneas las correspondientes a la misma variante de B)
- Si los dos factores son cualitativos, conviene ver los dos gráficos, posicionando uno u otro en el eje de abscisas
- Si un factor es cuantitativo conviene situarlo en el eje de abscisas

Ejemplos de posibles gráficos para la interacción entre dos factores A y B, con dos variantes cada uno









Interacciones con factor (o factores) cuantitativos: descomposición de la interacción A*B

- **Caso 1: A cualitativo y B cuantitativo**
  - Sea A (por ejemplo la Variedad) una factor cualitativo con I variantes (I-1 g.l.) y B (por ejemplo la Dosis de abonado) un factor cuantitativo con J niveles (J-1 g.l.)
  - El efecto simple de B se puede descomponer, como se ha visto, en I-1 componentes ortogonales: lineal, cuadrática, …
  - La interacción A*B (que tiene (I-1)x((J-1) g.l.) mide (en el ejemplo) si la respuesta al abonado difiere de unas variedades a otras, y se descompone en:
    - A * (efecto lineal de B) con I-1 g.l.
    - A * (efecto cuadrático de B) con I-1 g.l.
    - …
    - (Las fórmulas para calcular las sumas de cuadrados asociadas a estos términos se verán en los ejemplos)
  - La interacción A *(efecto lineal de B) analiza si el efecto lineal del factor B es diferente según la variante considerada de A
  - La interacción A *(efecto cuadrático de B) analiza si el efecto cuadrático del factor B es diferente según la variante de A

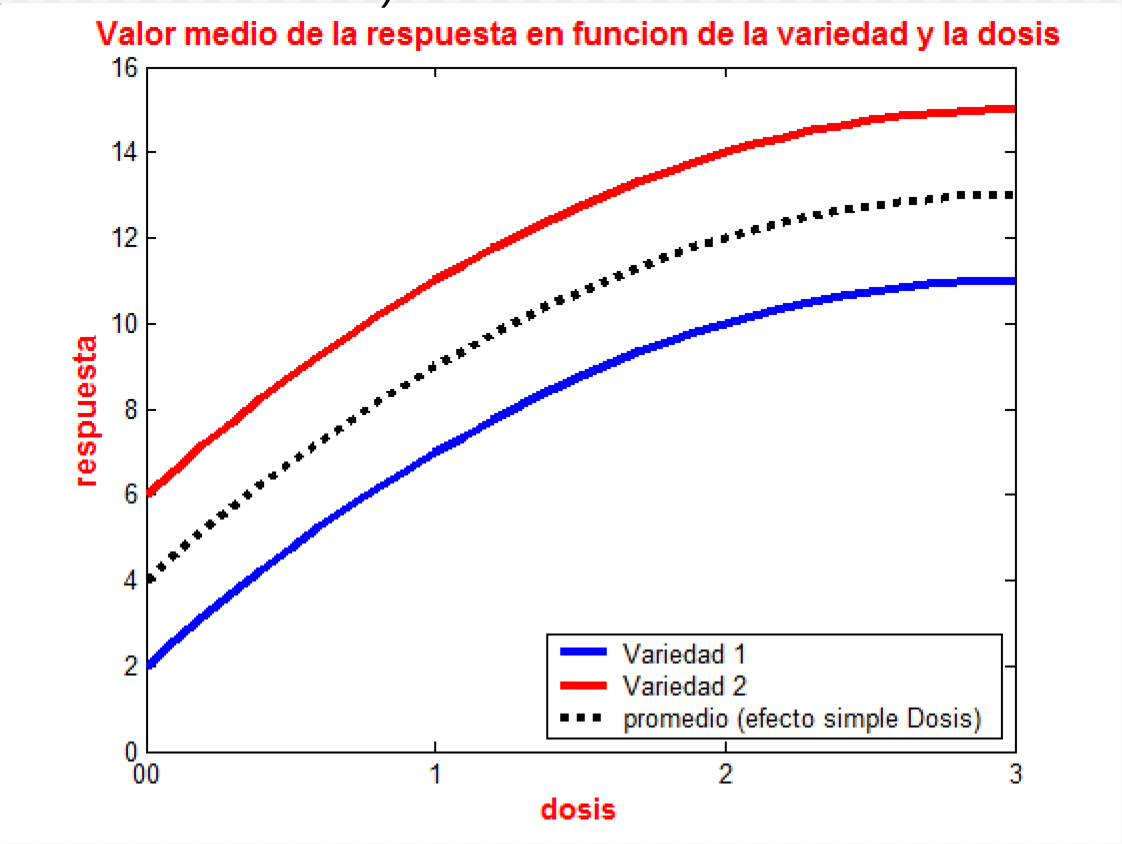
Variedad: factor cualitativo con 2 variantes
Dosis: factor cuantitativo con 4 niveles

Efecto simple del factor Variedad significativo

Componente lineal efecto simple Dosis significativa (positiva)

Componente cuadrática efecto simple Dosis significativa (negativa)

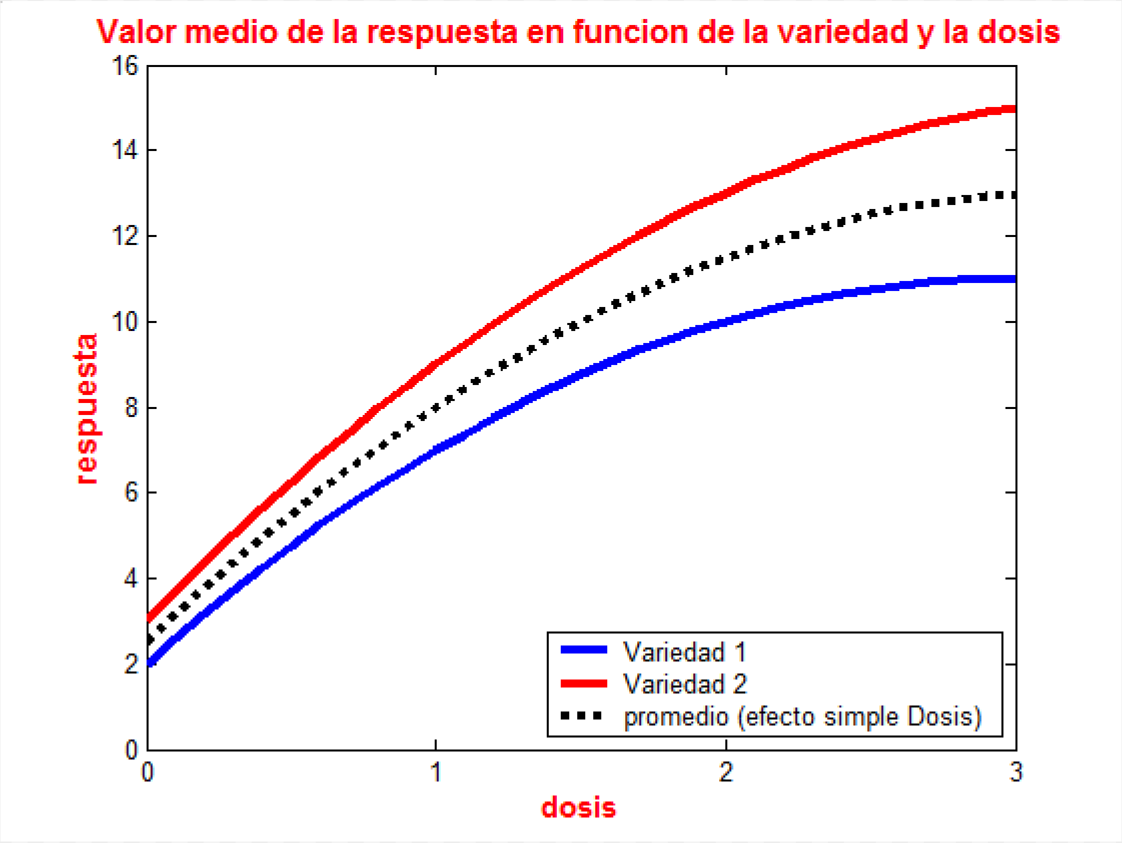
Interacción Variedad*Dosis no significativa (la forma de la función de respuesta es similar en ambas variedades)



Variedad: factor cualitativo con 2 variantes
Dosis: factor cuantitativo con 4 niveles

Interacción Variedad*(efecto lineal Dosis) significativa (el aumento de la respuesta al aumentar la dosis es mayor en la variedad 2 que en la 1)

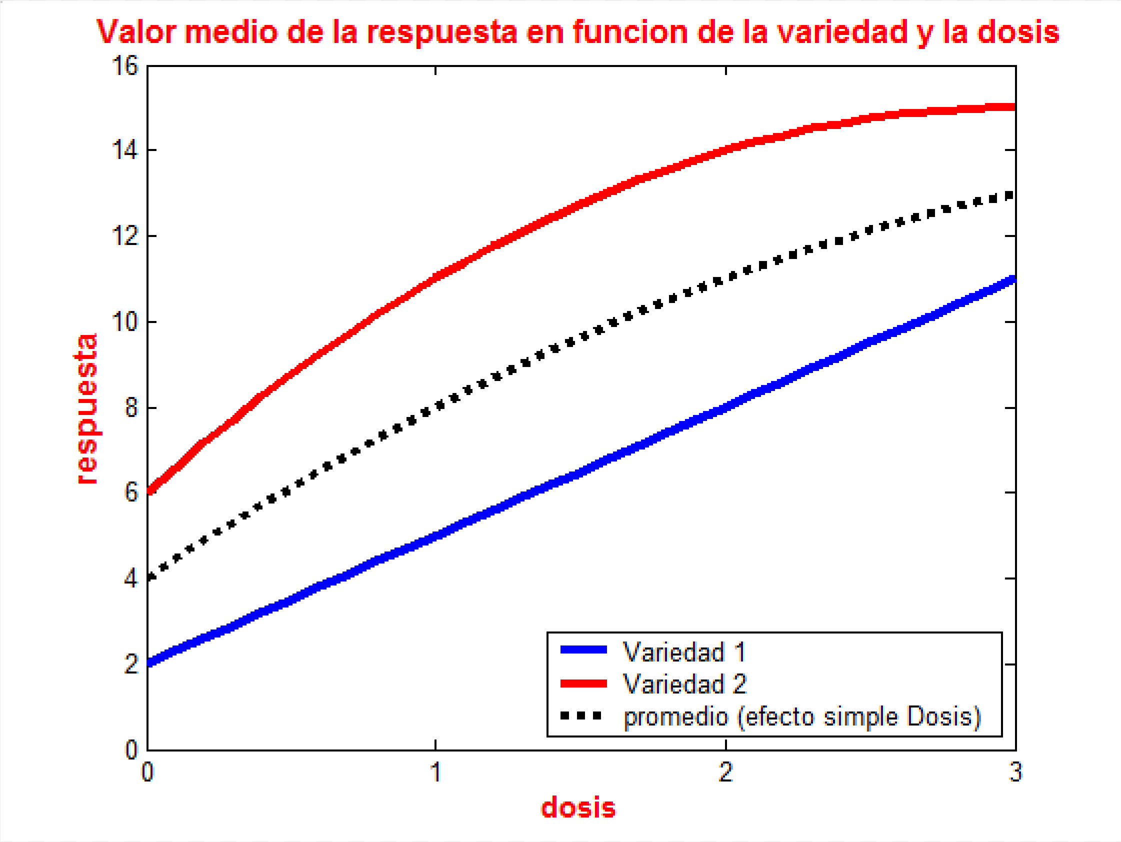
Interacción Variedad*(efecto cuadrático Dosis) no significativa (la curvatura de la función de respuesta es negativa y similar en ambas variedades)



Variedad: factor cualitativo con 2 variantes
Dosis: factor cuantitativo con 4 niveles

Interacción Variedad*(efecto lineal Dosis) no significativa (en promedio en la zona estudiada, el aumento de la respuesta al aumentar la dosis es similar en las dos variedades)

Interacción Variedad*(efecto cuadrático Dosis) significativa (la curvatura de la función de respuesta es diferente en la variedad 2 de la de la variedad 1)



Interacciones con factor (o factores) cuantitativos: descomposición de la interacción A*B

- **Caso 2: A cuantitativo y B cuantitativo**
  - Sea A (por ejemplo dosis de abonado nitrogenado) una factor cuantitativo con 3 niveles (2 g.l.). Llamamos $\mathbf{x_1}$ a los niveles de A
  - Sea B (por ejemplo dosis de abonado fosfórico) una factor cuantitativo con 3 niveles (2 g.l.). Llamamos $\mathbf{x_2}$ a los niveles de B
  - La relación que liga el valor medio de la respuesta **Y** (por ejemplo, el rendimiento) con los niveles de A y de B será, en general, una función **$E(Y) = f(x_1,x_2)$** que podrá aproximarse por un polinomio de un determinado grado.
  - La componente lineal del efecto simple de A corresponde al término x_1 y la componente cuadrática al término x_1^2 de ese polinomio
  - Análogamente, la componente lineal del efecto simple de B corresponde al término x_2 y la componente cuadrática al término x_2^2 de ese polinomio
  - ¿A qué corresponden los 4 grados de libertad de la interacción A*B?

Descomposición de los 4 grados de libertad de la interacción entre dos factores cuantitativos con 3 niveles

		Factor B	
		Comp. Lineal (término x_2)	Comp.Cuadrat. (término x_1^2)
Factor A	Comp. Lineal (término x_1)	Lineal x Lineal (término $x_1 x_2$)	Lineal x Cuad. (término $x_1 x_2^2$)
	Comp.Cuadrat. (término x_1^2)	Cuad. x Lineal (término $x_1^2 x_2$)	Cuad x Cad (término $x_1^2 x_2^2$)

En general la componente lineal x lineal "se lleva" la mayor parte de la Suma de Cuadrados de la interacción, por lo que puede resultar significativa aunque la interacción globalmente no lo sea

Realización de los cálculos para descomponer las interacciones dobles en componentes

- Statgraphics **no** incorpora opciones para decomponer en sus diferentes componentes la Suma de Cuadrados de una interacción doble en la que al menos uno de los factores implicados es de naturaleza cuantitativa
- Una exposición detallada de los cálculos a realizar con ese fin puede encontrarse en el Apartado 10.4.4 del documento *"capitulo_10_MEI.doc"* incluido en la documentación de este curso
- En algunos de los ejemplos que se desarrollan más adelante se verá de forma práctica la operativa para llevar a cabo las descomposiciones y las conclusiones que se derivan de las mismas

PREDICCIONES

- Tras la realización de un Anova, en algunas ocasiones (especialmente en contextos industriales) interesa obtener el valor medio previsto para una determinada combinación de variantes o niveles de los diferentes factores. (Por ejemplo, en un estudio con 3 factores si se seleccionan las variantes i de FI, j de FJ y k de FK, se trata de estimar la media m_{ijk} correspondiente a dicho tratamiento)
- La predicción se lleva a cabo adicionando a la media general obtenida los valores estimados de los efectos correspondientes que hayan resultado significativos. Así, si en el caso de 3 factores hubieran resultado significativos los 3 efectos simples y la interacción entre los dos primeros, la predicción sería:

$$m^*_{ijk} = \bar{x}_{..} + a_i + b_j + c_k + (ab)_{ij}$$

- donde: a_i es el efecto estimado de la variante i del factor FI (calculado como la diferencia entre la media de esta variante y la media general).
- b_j y c_k son los efectos estimados de las variantes j y k de los factores FJ y FK (calculados de forma análoga)
- y $(ab)_{ij}$ es el efecto estimado de la interacción entre la variante i de FI y la variante j de FJ, calculado mediante la expresión:

$$(ab)_{ij} = \bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$$

- Statgraphics permite calcular las predicciones restando a los datos los residuos estimados (ver siguiente diapositiva)

RESIDUOS

- Tras finalizar un Anova, a cada dato inicial se le asocia un **residuo**
- El residuo no es más que la diferencia entre el valor observado y el valor medio previsto para el tratamiento correspondiente
  - **Residuo = dato – predicción**
- El residuo recoge el efecto que sobre esa observación concreta han tenido todos los factores no controlados en la investigación
- Los residuos, analizados de forma gráfica, son una herramienta muy importante para estudiar la validez del modelo
- **Gráfico de los residuos en papel probabilístico normal**: permite detectar la presencia de datos anómalos y, en general, constatar la validez de la hipótesis de normalidad
- **Gráfico de los residuos frente a las predicciones**: permite analizar la validez de la hipótesis de homocedasticidad, así como posibles efectos no lineales no contemplados en el estudio
- **Gráfico de los residuos frente a las variantes de un factor**: permite detectar posible efectos sobre la varianza de dicho factor (ver siguiente diapositiva)

Análisis de efectos sobre la varianza

- El procedimiento visto en el tema anterior para investigar si un factor, al margen de su posible efecto sobre la media de la respuesta, tiene efectos sobre la varianza de la misma, se generaliza sin dificultad al caso multifactorial. (Ésta es una ventaja muy importante de este procedimiento)
- Tras realizar el Anova sobre los datos iniciales, obtener el residuo de cada observación (Statgraphics los obtiene directamente)
- Realizar un nuevo Anova utilizando como variable respuesta el cuadrado de los residuos, para ver cómo influyen los factores sobre la media de estos cuadrados (media que no es más que la varianza estimada en cada condición)
- Nota: La detección de efectos sobre varianzas es en general difícil, por lo que los resultados de este nuevo Anova deben considerarse con precaución si el número de grados de libertad del primer Anova es reducido

Ejemplo 1: Descomposición de una interacción cualitativo x cuantitativo

- En una experiencia para analizar la influencia de la variedad y de la altura de corte sobre el rendimiento de sorgo forrajero se ensayaron 4 alturas y 3 variedades, cultivándose 3 parcelas (elegidas al azar de un total de 36) para cada una de las 12 combinaciones posibles de variedad por altura. Los resultados obtenidos, en Tms. de forraje por Ha y año, se recogen en la tabla siguiente (datos en *sorgo.sgd*)

	VARIEDAD A	*VARIEDAD B*	*VARIEDAD C*
Altura 0.75	*68* *60* *62*	*52* *55* *61*	*66* *72* *68*
Altura 1	*91* *75* *86*	*62* *67* *60*	*83* *82* *78*
Altura 1.25	*90* *98* *94*	*64* *75* *74*	*72* *66* *74*
Altura 1.50	*105* *95* *99*	*68* *85* *83*	*61* *58* *58*

- a) Calcular el cuadro resumen del ANOVA y estudiar qué efectos son significativos.
- b) Descomponer el efecto de la altura en sus componentes lineal, cuadrática y cúbica y estudiar la significación de las mismas.
- c) Descomponer la interacción Variedad x Altura en sus 3 componentes (Variedad x lineal, Variedad x cuadrática y Variedad x cúbica) y estudiar la significación de cada una de ellas.
- d) ¿Afectan la variedad o la Altura a la varianza del rendimiento?

Ejemplo 2: Optimización de un proceso industrial. PFE con 4 factores a dos niveles (Plan 2^4)

- Se realizó un experimento para mejorar la fuerza de adhesión (RESIST) obtenida en el proceso de adhesivado de planchas de poliuretano utilizadas en el revestimiento interior de diversos equipos. El objetivo perseguido era garantizar una resistencia mínima de 4 newtons. El experimento fue un PFE con 4 factores a 2 niveles (resultados en el archivo *aditivpe.sgd*)
- Realizar el Anova de los resultados incluyendo todos los efectos simples e interacciones dobles. Repetir el análisis incluyendo en la SCresid las interacciones claramente no significativas
- Interpretar los resultados obtenidos indicando qué efectos son significativos. Intentar hallar una interpretación técnica a la interacción que ha resultado significativa. Obtener cuáles serían sus niveles operativos óptimos para maximizar la resistencia media obtenida.
- ¿Qué resistencia media cabe esperar operando en las condiciones óptimas halladas en el experimento?
- Estudiar si existe algún efecto significativo de los factores sobre la varianza de la resistencia obtenida. Interpretar técnicamente la conclusión hallada. ¿Cuál es la varianza previsible trabajando en las condiciones óptimas propuestas?
- ¿Cuál es la probabilidad, operando en estas condiciones, de obtener una plancha con una resistencia inferior a 4 newtons?
- Si se considera suficiente que la probabilidad anterior no supere el 1%, ¿qué condiciones operativas propondrías para el proceso, con el fin de que, cumpliendo el requisito técnico exigido, resultara más económico?

Ejemplo 3: Efecto de la temperatura y el sexo en la longevidad de adultos de *aganaspis daci*

- El diseño utilizado ha sido un Plan Factorial Equilibrado con 3 réplicas para cada una de las 10 combinaciones entre los dos factores estudiados:
  - Temperatura: factor cuantitativo con 5 niveles equiespaciados entre 15ºC y 35ºC
  - Sexo: factor cualitativo con dos variantes (machos y hembras)
- Cada unidad experimental ha sido un cilindro con 10 individuos, siendo la respuesta constatada en la misma el valor medio de las longevidades (en días) de dichos individuos.
- Los resultados del experimento están en *aganaspis.sf3*
- Realizar un Anova de los resultados, descomponiendo el efecto simple de la temperatura y su interacción con sexo en sus componentes lineal, cuadrática y resto
- Estudiar si el sexo o la temperatura influyen sobre la varianza de la longevidad