

Análisis avanzado de datos en Ingeniería Informática



Rosa M^a Alcover Arándiga

Departamento de Estadística e Investigación Operativa
Aplicadas y Calidad



Objetivo de la asignatura

La asignatura pretende proporcionar al alumno los conocimientos necesarios para trabajar los planteamientos e ideas de las modernas estrategias de la **Calidad Total**, con especial relevancia en la utilización de las herramientas estadísticas orientadas a la mejora de la calidad de los productos y de la productividad de los procesos. Todo ello en el contexto de la ingeniería informática

Actividades desarrolladas en la ***Fase de diseño de productos y procesos***

Calidad Competitiva

Identificar y **cuantificar** los ***efectos*** que los parámetros de los sistemas tienen sobre las características de rendimiento y calidad

- ▶ Obtención de **configuraciones óptimas**
- ▶ Identificar **causas de problemas**
- ▶ Seleccionar la **mejor alternativa** para resolverlos

Algunos ejemplos de aplicación

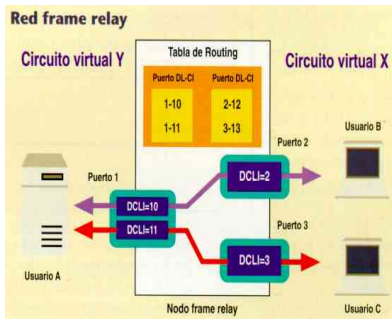


¿Cuál será la “**mejor**” **configuración** del circuito de comunicaciones de los nodos de la red con el fin de que el tiempo de transmisión sea el menor posible?



¿Qué sistema de los considerados es **significativamente más rápido** en la ejecución de una serie de aplicaciones?

Algunos ejemplos de aplicación



¿Qué parámetro es responsable de que el circuito de comunicaciones del sistema sea de *baja calidad*?



¿A partir de **qué carga** debo trabajar con el sistema **X** o con el sistema **Y** para que el tiempo de respuesta sea lo más bajo posible?

Algunos ejemplos de aplicación



Un ingeniero de software quiere determinar si el sistema operativo influye en el rendimiento de cierta aplicación que se está diseñando. Para ello prueba distintas distribuciones de Linux y diferentes Windows, ¿podría decirse que **afecta el sistema operativo al rendimiento** de esta aplicación?



Un administrador de red está interesado en comprobar si el servidor de páginas *web* influye en el tiempo de respuesta a distintas peticiones. Para ello instala tres servidores diferentes en la misma máquina y, desde distintos puestos, encarga a varios usuarios de la *web* que visiten diferentes páginas en un orden totalmente aleatorio. ¿**Afecta el factor servidor web al tiempo de respuesta** de la red?

Contenidos de la asignatura



Calidad Total



Herramientas de mejora



Técnicas estadísticas

Ejemplos

Ejemplos



Diseño de Experimentos

ANOVA

Fracciones Factoriales
Orthogonal arrays

Modelos de Regresión

Unidades didácticas

UD 1	Panorámica general de la Calidad Total. Herramientas estadísticas de la Calidad Total. Conceptos básicos del Control Estadístico de Procesos. Control <i>on-line</i> frente a control <i>off-line</i>
UD 2	Diseño de Experimentos. Planes factoriales. Análisis de datos procedentes de un diseño experimental. ANOVA. Gráfico de Daniel
UD 3	Fracciones Factoriales. Introducción al Diseño Robusto de productos y procesos. <i>Orthogonal Arrays</i>
UD 4	Modelos de Regresión. Generalizaciones del modelo de Regresión Lineal Simple. Predicciones. Validación del modelo. Análisis de Residuos
Prácticas	Programa de prácticas: Cada unidad didáctica lleva asociada la realización de prácticas de laboratorio utilizando software estadístico <i>Statgraphics Centurion</i> .

Ejemplo de análisis de datos procedentes de un diseño fraccionado L_{18}

STATGRAPHICS Centurion - StatFolio sin título

Archivo Editar Graficar Describir Comparar Relacionar Pronósticos CEP DDE SnapStats!! Herramientas Ver Ventana Ayuda

Libro de Datos
StatAdvisor
StatGallery
StatReporter
Comentarios del StatFolio
Atributos del Diseño de C

StickybitsL18.sf3

	Unidad_Disco	Tabla	INODE	Tam_Memoria	Sticky_bits	Tpo_Respuesta
1	1	1	1	1	1	3,65
2	1	1	2	2	2	5,28
3	1	1	3	3	3	2,06
4	1	2	1	1	2	3,53
5	1	2	2	2	3	2,26
6	1	2	3	3	1	4,55
7	1	3	1	1	1	2,37
8	1	3	2	2	2	5,62
9	1	3	3	3	3	3,87
10	2	1	1	1	3	2,13
11	2	1	2	2	1	4,08
12	2	1	3	3	2	4,45
13	2	2	1	1	3	1,81
14	2	2	2	2	1	5,87
15	2	2	3	3	2	3,42
16	2	3	1	1	2	2,66
17	2	3	2	2	3	2,92
18	2	3	3	3	1	4,42

Ejemplo de análisis de datos procedentes de un diseño fraccionado L_{18}

Asistente de Diseño de Experimentos

1) Definir respuestas 3) Seleccionar diseño 5) Seleccionar ejecuciones 7) Guardar experimento 9) Optimizar respuestas 11) Aumentar diseño
 2) Definir factores exp. 4) Especificar modelo 6) Evaluar diseño 8) Analizar datos 10) Guardar resultados 12) Extrapolación

Asistente de diseño de experimentos - Definir Factores

Archivo de diseño: <sin título>
 Comentario: Tiempo de Respuesta sistema UNIX

Número de factores controlables del proceso: 4 Número de componentes controlables de mezcla: 0 Número de factores de ruido: 0

Factor	Nombre	Unidades	Tipo	Función	Bajo	Alto	Niveles
A	Unidad_disco		Catagórico	Controlable	-1,0	1,0	1,2
B	Tabla INODE		Catagórico	Controlable	-1,0	1,0	1,2,3
C	Tam_memoria		Catagórico	Controlable	-1,0	1,0	1,2,3
D	Sticky_bits		Catagórico	Controlable	-1,0	1,0	1,2,3
E	Factor_E		Continuo		-1,0	1,0	1,2,3,4
F	Factor_F		Continuo		-1,0	1,0	1,2,3,4
G	Factor_G		Continuo		-1,0	1,0	1,2,3,4
H	Factor_H		Continuo		-1,0	1,0	1,2,3,4
I	Factor_I		Continuo		-1,0	1,0	1,2,3,4
J	Factor_J		Continuo		-1,0	1,0	1,2,3,4
K	Factor_K		Continuo		-1,0	1,0	1,2,3,4
L	Factor_L		Continuo		-1,0	1,0	1,2,3,4
M	Factor_M		Continuo		-1,0	1,0	1,2,3,4

Total para componentes controlables de mezcla: 1,0 Factores A-M Factores N-Z

Aceptar Atrás Cancelar Ayuda

Asistente de Diseño de Experimentos

1) Definir respuestas 3) Seleccionar diseño 5) Seleccionar ejecuciones 7) Guardar experimento 9) Optimizar respuestas 11) Aumentar diseño
 2) Definir factores exp. 4) Especificar modelo 6) Evaluar diseño 8) Analizar datos 10) Guardar resultados 12) Extrapolación

Asistente de Diseño de Experimentos

Paso 1: Definir la
 Nombre

Paso 2: Definir lo
 Nombre

Paso 3: Seleccionar
 Se mostrarán una

Asistente de diseño de experimentos - Seleccionar diseño

Archivo de diseño: <sin título>
 Comentario: Tiempo de Respuesta sistema UNIX

Diseño paramétrico robusto
 Vector combinad
 Vector cruzado

Opciones...	Segmento	Factores	Ejecución	Bloques	Diseño
Opciones...	Factores de proceso	4	0	0	Diseño especificado por el usuario
Opciones...	Componentes de mezcla	0	0	0	
Opciones...		0	0	0	
	COMBINADO	4	0	0	Muestras por ejecución: 1

BLOQUE Unidad_disco Tabla INODE Tam_memoria Sticky_bits

Opciones de diseño multifactorial categórico

Tipo de diseño
 Factorial
 Componentes de la varianza (Jerárquic.
 Cuadrado Latino
 Cuadrado Greco-Latino
 Cuadrado Hiper-Greco-Latino
 Usuario-especificado

Diseño replicado
 Número: 0

Aleatorizar

Aceptar Cancelar Ayuda

Aceptar Cancelar Aleatorizar Ayuda

Ejemplo de análisis de datos procedentes de un diseño fraccionado L_{18}

ANOVA Multifactorial

Variable Dependiente: Tpo_Respuesta

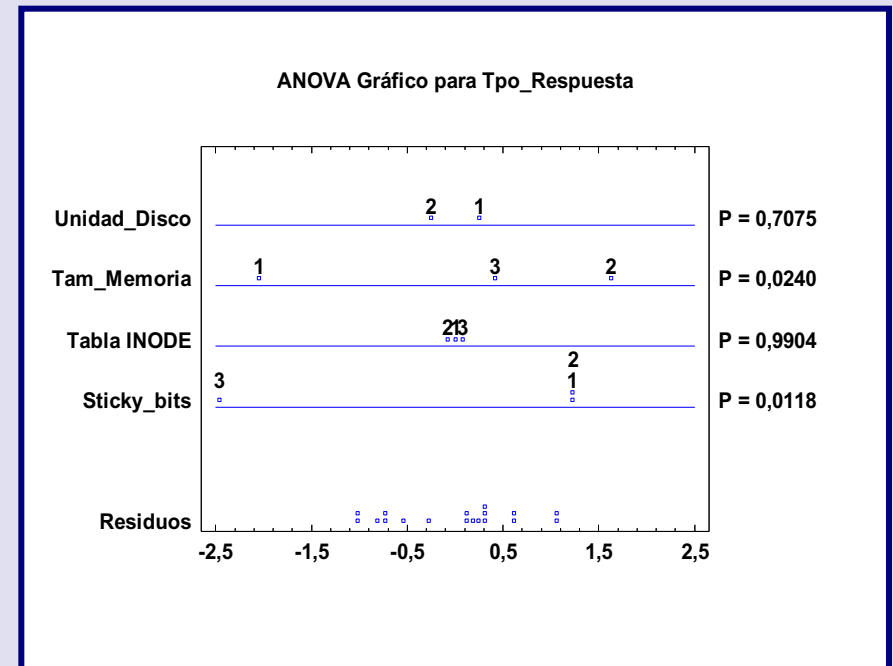
Factores: Sticky_bits, Tabla INODE, Tam_Memoria, Unidad_Disco

Covariables:

(Selección:)

Ordenar nombres de columna

Aceptar Cancelar Borrar Transformar... Ayuda



Análisis de Varianza para Tpo_Respuesta - Suma de Cuadrados Tipo III

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
EFFECTOS PRINCIPALES					
A:Sticky_bits	10,89	2	5,44502	7,15	0,0118
B:Tabla INODE	0,0147	2	0,00735	0,01	0,9904
C:Tam_Memoria	8,44813	2	4,22407	5,54	0,0240
D:Unidad_Disco	0,113606	1	0,113606	0,15	0,7075
RESIDUOS	7,61958	10	0,761958		
TOTAL (CORREGIDO)	27,0861	17			

Ejemplo de análisis de datos procedentes de un diseño fraccionado L_{18}

Tablas y Gráficos

TABLAS

- Resumen del Análisis
- Tabla ANOVA
- Tabla de Medias
- Pruebas de Múltiple Rangos

GRÁFICOS

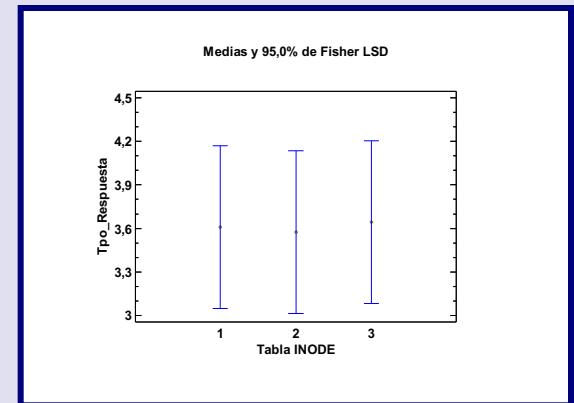
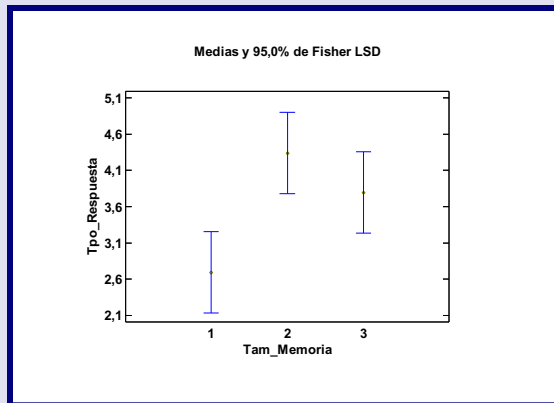
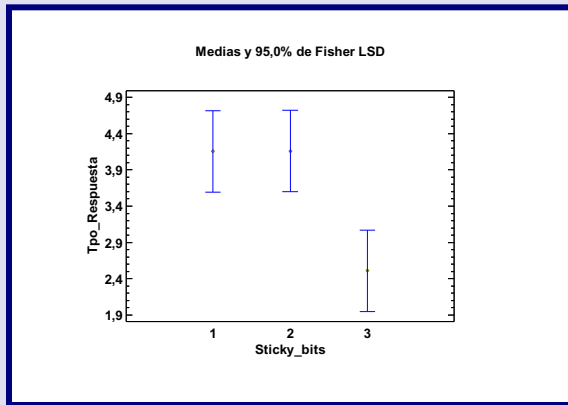
- Gráfico de Dispersión
- ANOVA Gráfico
- Gráfico de Medias
- Gráfico de Interacción
- Residuos vs Niveles de Factor
- Residuos vs Predichos
- Residuos vs Número de Fila

Aceptar
 Cancelar
 Todos
 Almacén
 Ayuda

Tabla de Medias por Mínimos Cuadrados para Tpo_Respuesta con intervalos de confianza del 95,0%

			Error	Lmite	Lmite
Nivel	Casos	Media	Est.	Inferior	Superior
MEDIA GLOBAL	18	3,60833			
Sticky_bits					
1	6	4,15667	0,356361	3,36264	4,95069
2	6	4,16	0,356361	3,36598	4,95402
3	6	2,50833	0,356361	1,71431	3,30236
Tabla INODE					
1	6	3,60833	0,356361	2,81431	4,40236
2	6	3,57333	0,356361	2,77931	4,36736
3	6	3,64333	0,356361	2,84931	4,43736
Tam_Memoria					
1	6	2,69167	0,356361	1,89764	3,48569
2	6	4,33833	0,356361	3,54431	5,13236
3	6	3,795	0,356361	3,00098	4,58902
Unidad_Disco					
1	9	3,68778	0,290967	3,03946	4,33609
2	9	3,52889	0,290967	2,88057	4,17721

Ejemplo de análisis de datos procedentes de un diseño fraccionado L_{18}



Obtención de la **configuración óptima** de los cuatro parámetros analizados para **minimizar el tiempo de respuesta** del sistema