ANÁLISIS DE RESULTADOS Y DISEÑO DE EXPERIMENTOS

INTRODUCCIÓN

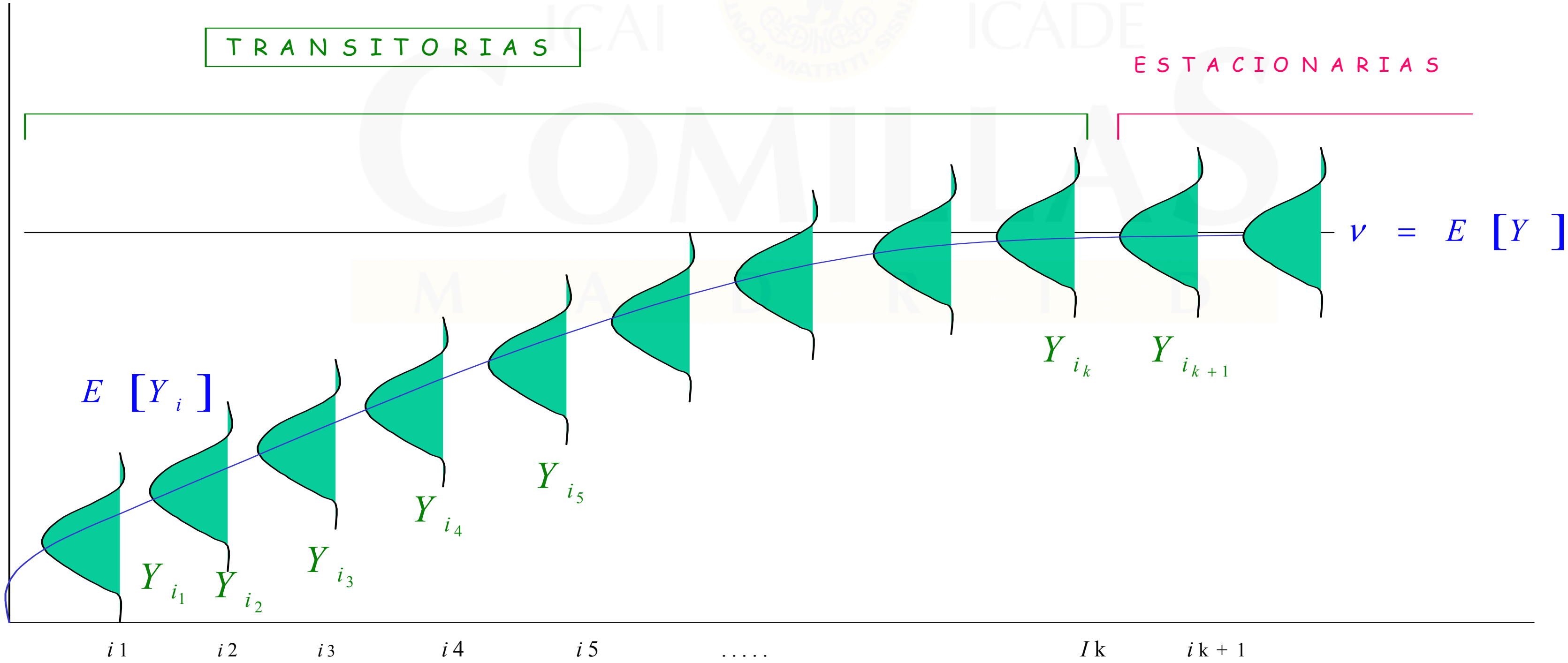
- UN ESTUDIO DE SIMULACIÓN BUSCA RESPUESTAS A PREGUNTAS SOBRE EL SISTEMA A TRAVÉS DE LA INFORMACIÓN QUE PROPORCIONAN LOS EXPERIMENTOS CON EL MODELO DEL SISTEMA
- LOS EXPERIMENTOS RESPONDEN A PREGUNTAS: ¿QUÉ PASARÍA SÍ? (WHAT-IF)
- LAS RESPUESTAS SERVIRÁN DE SOPORTE A UNA DECISIÓN RACIONAL SOBRE EL SISTEMA → INTERESA SEAN EXPRESADAS NUMÉRICAMENTE PARA CADA ALTERNATIVA
- LAS ALTERNATIVAS CONSTITUIRÁN UNA VARIANTE DEL MODELO O ESCENARIO DE SIMULACIÓN CON LAS QUE REALIZAREMOS LOS EXPERIMENTOS → ESTIMACIÓN DE VARIABLES RESPUESTA → ESTADÍSTICA:

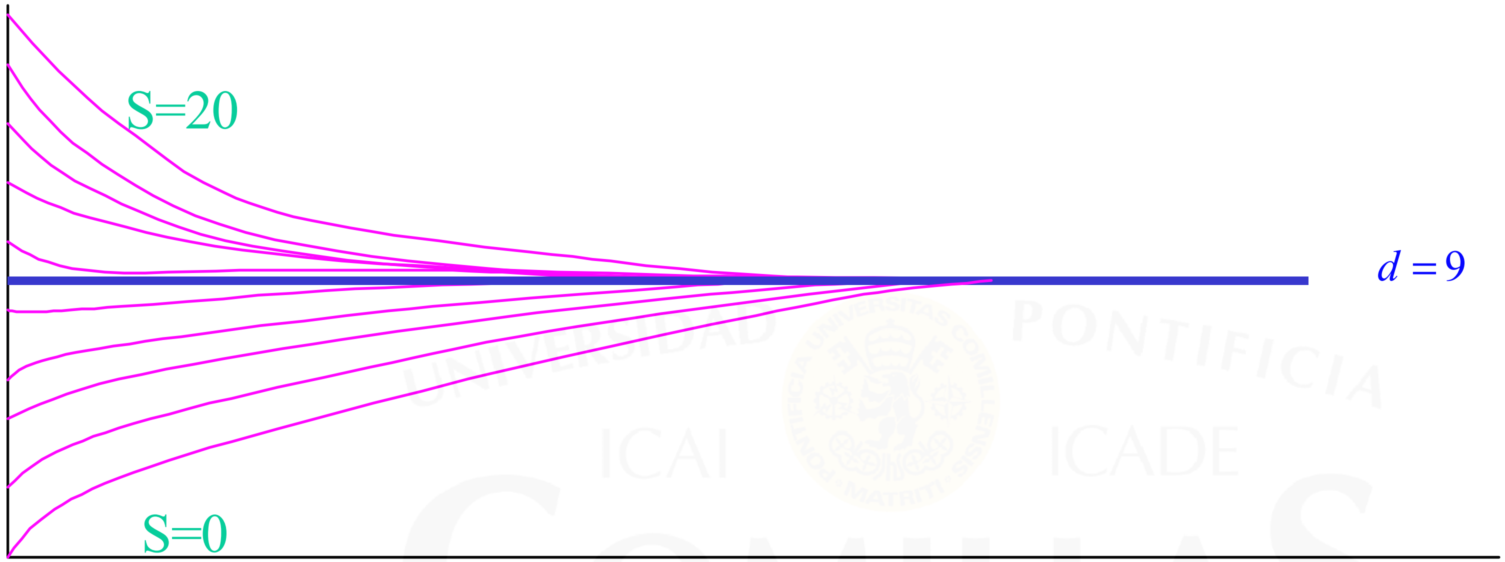
- MUESTREO - REDUCCIÓN DE LA VARIANZA - ESTIMACIÓN - DISEÑO EXPERIMENTOS

1. COMPORTAMIENTO TRANSITORIO Y ESTACIONARIO DE UN PROCESO ESTOCÁSTICO

- **I: CONDICIONES INICIALES.**
- $F_i(y/I)$ **: DISTRIBUCIÓN TRANSITORIA EN INSTANTE** i **COND. INICI. I.**

SI $F_i(y/I) \xrightarrow[i\to\infty]{} F(y) \quad \forall y, \forall I$ **,** $F(y)$ **: DISTRIBUCIÓN ESTACIONARIA**





S: nº clientes en sistema en t=0 $E[D_i / S]$: nº medio clientes en sistema

2. TIPOS DE SIMULACIÓN SEGÚN ANÁLISIS DE RESULTADOS

- SIMULACIÓN CON HORIZONTE FINITO: EXISTE UN EVENTO "NATURAL" E QUE ESPECIFICA LA LONGITUD DE CADA REPLICACIÓN. EN ESE EVENTO EL SISTEMA SE REINICIALIZA, OBTENIENDO M.A.S. LAS CONDICIONES INICIALES GENERALMENTE AFECTAN A LAS MEDIDAS DE DESARROLLO, HAN DE SER REPRESENTATIVAS DEL SISTEMA REAL → PERIODO DE "CALENTAMIENTO O ARRANQUE" (WARM UP) Ó ALEATORIZACIÓN CONDICIONES INICIALES.
- SIMULACIÓN CON HORIZONTE INFINITO: NO EXISTE TAL EVENTO QUE INDIQUE EL FINAL DE LA REPLICACIÓN. POSIBILIDADES:

A) EXISTE DISTRIBUCIÓN ESTACIONARIA → ESTIMAR PARÁMETROS ESTACIONARIOS

B) NO ESTACIONARIA, SÍ POR CICLOS → PARÁMETROS ESTACIONARIOS DEL CICLO.

C) NO ESTACIONARIA, PUES LOS DATOS DE ENTRADA VARÍAN EN EL TIEMPO → CONSIDERAR QUE CADA VEZ QUE CAMBIAN ES UN FINAL DE HORIZONTE

3. ESTIMACIÓN VARIABLES RESPUESTA: ESTIMACIÓN MEDIAS (ESPERANZAS)

EN GENERAL, VALOR ESPERADO DE VARIABLE RESPUESTA SE ESTIMA MEDIANTE MEDIA MUESTRAL DE LAS OBSERVACIONES. TAMBIÉN INTERVALO DE CONFIANZA (PRECISIÓN).

$$Y_1,\ldots,Y_n \;\rightarrow\; \overline{Y}=\frac{\sum_{i=1}^{n} Y_i}{n} \qquad S^2=\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}$$

PARA UN NIVEL DE CONFIANZA α,

$$\boxed{\overline{Y}\pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}}$$

(DE 100 INTERVALOS CONFIAMOS EN QUE EN AL MENOS α. 100 ESTARÁ LA MEDIA)

- MUESTREO DE DIMENSIÓN FIJA: n FIJADO ANTEMANO $\rightarrow$ PRECISIÓN LA QUE RESULTE
- MUESTREO SECUENCIAL: PRECISIÓN FIJADA ANTEMANO (ANCHURA DEL INTERVALO) $\rightarrow$ INDETERMINADO TAMAÑO DE MUESTRA (FIJAR Y SI NO SE ALCANZA PRECISIÓN, SEGUIR)

4. ESTIMACIÓN DE PARÁMETROS ESTACIONARIOS: EL PROBLEMA DEL ESTADO INICIAL TRANSITORIO

- SI LA SIMULACIÓN ES DE HORIZONTE FINITO, EL PERIODO TRANSITORIO HA DE TENERSE EN CUENTA. (INCLUSO PUEDE SER EL OBJETIVO ESTUDIAR SU COMPORTAMIENTO).
- SI SE PRETENDE ESTIMAR UN PARÁMETRO ESTACIONARIO O DE COMPORTAMIENTO "NORMAL", ES DECIR, $\nu = \lim_{i\to\infty} E\left[Y_i\right]$, HAY QUE EVITAR LAS INFLUENCIAS DEL ESTADO INICIAL.

MÉTODOS ESTIMADORES PUNTUALES Y POR INTERVALOS PARA MEDIA ESTACIONARIA

A) REPLICACIÓN/ELIMINACIÓN

B) PROCEDIMIENTO POR LOTES

C) PROCEDIMIENTOS REGENERATIVOS

A) REPLICACIÓN/ELIMINACIÓN:

n REPLICACIONES INDEPENDIENTES LONGITUD m. DETERMINAR PERIODO ARRANQUE DE LONGITUD l $(l << m)$, Y SE ELIMINAN ESAS OBSERVACIONES EN LA ESTIMACIÓN:

$$Y_{11},\ldots,Y_{1m},\ldots,Y_{n1},\ldots,Y_{nm} \rightarrow X_1 = \frac{\sum_{i=l+1}^{m} Y_{1i}}{m-l}, \ldots, X_n = \frac{\sum_{i=l+1}^{m} Y_{ni}}{m-l} \rightarrow \overline{X} \pm t_{n-1,\alpha/2} \frac{S_X}{\sqrt{n}}$$

PROBLEMA: DESVIACIÓN RESPECTO ν. DIFICULTAD: ELECCIÓN PERIODO TRANSITORIO

B) PROCEDIMIENTO POR LOTES:

UNA ÚNICA REPLICACIÓN (SÓLO UN PERIODO DE ARRANQUE QUE DEBE SER ELIMINADO). EL RESTO SE DIVIDE EN n LOTES DE TAMAÑO k. k Y n "SUFICIENTEMENTE" GRANDES, LOS BLOQUES PUEDAN CONSIDERARSE INDEPENDIENTES Y MEDIAS NORMALES (TCL)

$$\overline{Y}(n,k) = \frac{\sum_{j=1}^{n} \overline{Y}_j(k)}{n} = \frac{\sum_{i=1}^{nk} Y_i}{nk} \qquad \underbrace{Y_1,\ldots,Y_k}_{\overline{Y}_1(k)}, \underbrace{Y_{k+1},\ldots,Y_{2k}}_{\overline{Y}_2(k)},\ldots \underbrace{Y_{(n-1)k+1},\ldots,Y_{nk}}_{\overline{Y}_n(k)} \qquad S^2(n,k) = \frac{\sum_{j=1}^{n}(\overline{Y}_j(k) - \overline{Y}(n,k))^2}{n-1}$$

PROBLEMA: ESTIMACIÓN BAJA $\widehat{Var(\hat{\nu})}$ DIFICULTAD: TAMAÑO k PARA NO CORRELACIÓN

C) PROCEDIMIENTOS REGENERATIVOS:

UNA ÚNICA REPLICACIÓN, COMO ANTES, PERO LOS BLOQUES NO SON DEL MISMO TAMAÑO, SINO QUE SE TOMA UN PUNTO DE REGENERACIÓN PARA DETERMINAR DÓNDE ACABA, SIENDO VARIABLE LA LONGITUD (N_i):

$$Y_1, Y_2 \dots, Y_{B_1}, \; Y_{B_1+1}, \dots, Y_{B_2}, \; \dots, \; Y_{B_{n-1}+1}, \dots, Y_{B_n} \qquad X_j = \sum_{i=B_{j-1}+1}^{B_j} Y_i$$

$$N_1 = B_1, \; N_2 = B_2 - B_1, \; N_n = B_n - B_{n-1} \qquad N_j = B_j - B_{j-1} + 1$$

$$\left.\begin{array}{l} X_j \to \bar{X}, S_X^2 \\ N_j \to \bar{N}, S_N^2 \end{array}\right\} \quad \bar{Z} = \frac{\bar{X}}{\bar{N}} \qquad S_{X,N}^2 = \frac{\sum_{j=1}^{n} (X_j - \bar{X})(N_j - \bar{N})}{n-1}; \quad S^2 = S_X^2 - 2\bar{Z} S_{X,N}^2 + \bar{Z}^2 S_N^2$$

$$\boxed{\bar{Z} \pm z_{\alpha/2} \frac{S}{\bar{N}\sqrt{n}}}$$

PROBLEMA: SESGO (INFRAESTIMACIÓN) DE $\widehat{Var(\hat{v})}$

DIFICULTAD: NO HAY PUNTO DE REGENERACIÓN.

- TAMAÑOS N_i DEMASIADO PEQUEÑOS (NO INDEPENDENCIA)
- TAMAÑOS N_i DEMASIADO GRANDES (n PEQUEÑO)

5. TÉCNICAS DE REDUCCIÓN DE LA VARIANZA.

OBJETIVO: INCREMENTAR LA EFICIENCIA ESTADÍSTICA DEL ANÁLISIS DE SIMULACIÓN.

MEDIANTE → REDUCCIÓN DE LA VARIANZA DE LAS VARIABLES DE SALIDA (SIN MODIFICAR LA MEDIA). SE OBTIENE ASÍ MAYOR PRECISIÓN (INTERVALOS DE CONFIANZA MÁS AJUSTADOS) PARA EL MISMO NÚMERO DE DATOS, O LA PRECISIÓN DESEADA CON MENOS PASADAS DE SIMULACIÓN.

OBSERVACIONES:

- LOS MÉTODOS DEPENDEN DEL MODELO EN ESTUDIO.
- NORMALMENTE NO ES POSIBLE SABER DE ANTEMANO CUÁNTO SE VA A PODER REDUCIR LA VARIANZA (O SI SE VA A PODER REDUCIR).
- ALGUNAS TÉCNICAS PUEDEN AUMENTAR EL COSTE COMPUTACIONAL → BUSCAR EQUILIBRIO.

TÉCNICAS DE REDUCCIÓN DE LA VARIANZA:

- ***MUESTREO CORRELADO.***
- ***VARIABLES DE CONTROL.***
- ***VARIABLES ANTITÉTICAS.***
- ***CONDICIONAMIENTO***
- ***MUESTREO ESTRATIFICADO.***
- ***MUESTREO POR IMPORTANCIA***

6.1 MUESTREO CORRELADO (NÚMEROS ALEATORIOS COMUNES)

OBJETIVO: COMPARAR DOS O MÁS CONFIGURACIONES ALTERNATIVAS PARA EL SISTEMA "BAJO CONDICIONES DE EXPERIMENTACIÓN SIMILARES".

IDEA BÁSICA: X_{1J}, X_{2J} OBSERVACIONES PARA LA 1ª Y 2ª CONFIGURACIONES EN TIEMPO J

QUEREMOS ESTIMAR $\xi = \mu_1 - \mu_2 = E(X_{1J}) - E(X_{2J})$

SEA $Z_J = X_{1J} - X_{2J}$, CON $E(Z_J) = \xi$

ASÍ, $\overline{Z}(N) = \dfrac{\sum_{1}^{n} Z_j}{n}$ ES UN ESTIMADOR INSESGADO DE ξ.

$$V[\overline{Z}(N)] = \frac{V(Z_j)}{n} = \frac{V(X_{1j}) + V(X_{2j}) - 2\mathrm{Cov}(X_{1j}, X_{2j})}{n}$$

METODOLOGÍA: UTILIZAR LOS MISMOS VALORES U(0,1) PARA SIMULAR CADA UNA DE LAS CONFIGURACIONES A LO LARGO DEL TIEMPO. ASÍ, $\mathrm{Cov}(X_{1j}, X_{2j}) > 0$.

PROBLEMA: ¡CUIDADO CON LA SINCRONIZACIÓN!

6.2 VARIABLES ANTITÉTICAS

OBJETIVO: INDUCIR UNA CORRELACIÓN NEGATIVA ENTRE LAS SUCESIVAS SIMULACIONES DE UNA MISMA CONFIGURACIÓN, PARA CONSEGUIR REDUCIR LA VARIANZA.

IDEA BÁSICA:

X_1, X_2 ESTIMACIONES DE LA VARIABLE RESPUESTA CON DOS SIMULACIONES.

$\overline{X} = \frac{X_1 + X_2}{2}$ $\quad V[\overline{X}]=1/4(V(X_1)+V(X_2)+2Cov(X_1,X_2))$

SI X_1, X_2 INDEPEND. $Cov(X_1,X_2) = 0$; MEJOR SI $Cov(X_1,X_2) < 0$.

SI $U_1,\ldots U_N$ SON $U(0,1)$; $1-U_1 \ldots, 1-U_N$ SON $U(0,1)$

$X_1,\ldots,X_N$ $\qquad X'_1 \quad X'_N$

ENTONCES X_J Y X'_J ESTÁN CORRELADAS NEGATIVAMENTE.

APLICACIÓN: HACER UNA SIMULACIÓN CON $U_1,\ldots U_N$ Y LA SIGUIENTE CON $1-U_1 \ldots, 1-U_N$.

PROBLEMA: AUNQUE LAS VARIABLES DE ENTRADA ESTÉN CORRELADAS NEGATIVAMENTE, SI EL SISTEMA ES COMPLEJO LAS DE SALIDA PUEDEN NO ESTARLO.

6.3 VARIABLES DE CONTROL

OBJETIVO: USAR CORRELACIÓN ENTRE CIERTAS VARIABLES PARA REDUCIR VARIANZA

IDEA BÁSICA: X: V.A. DE SALIDA (EJ.: TIEMPO MEDIO DE ESPERA EN COLA DE LOS PRIMEROS 100 CLIENTES). SE DESEA ESTIMAR $\mu=E(X)$.

Y: V.A. QUE APARECE EN EL PROCESO DE SIMULACIÓN, CORRELADA CON X (POSITIVA O NEGATIVAMENTE), Y CON ESPERANZA $\nu = E(Y)$ CONOCIDA.

(EJ: MEDIA DE TIEMPOS DE SERVICIO DE PRIMEROS 99 CLIENTES QUE ACABAN SERVICIO)

VALOR OBSERVADO PARA Y ($Y>\nu$, $Y<\nu$) PERMITE AJUSTAR X. Y ES VARIABLE DE CONTROL PARA X

OBSERVACIÓN: NO DEPENDE DEL SIGNO DE LA CORRELACIÓN ENTRE X E Y.

CUANTIFICAR AJUSTE DE LA VARIABLE X (CON DESVIACIÓN DE Y DE SU MEDIA, $Y-\nu$). SE CONSIDERA EL ESTIMADOR DE CONTROL: $X_C = X - a(Y-\nu)$. (EST. INSESGADO DE μ)

$V[X_C]=V[X]+a^2V[Y]-2a\,Cov[X,Y] \Rightarrow X_C$ MENOR VARIANZA QUE $X \Leftrightarrow 2a\,Cov[X,Y]>a^2V[Y]$

PROBLEMA: DETERMINAR Y Y A PARA QUE ESTO SEA CIERTO.

6.4 CONDICIONAMIENTO

OBJETIVO: ELIMINAR PARTE DE LA VARIABILIDAD DEL MODELO SUSTITUYENDO ESTIMACIONES POR VALORES EXACTOS CONOCIDOS

IDEA BÁSICA: X: VARIABLE DE SALIDA E[X]= μ (ESTIMAR)

EXISTE OTRA VARIABLE ALEATORIA Z POSIBLE CALCULAR ANALÍTICAMENTE $E[X/Z=z]$

ESTIMADOR POR CONDICIONAMIENTO (INSESGADO): $\mu = E[X] = E_Z[E(X/Z)]$.

SI Z DISCRETA DISTRIBUCIÓN DESCONOCIDA: $E_Z[E(X/Z)] = \sum_z E(X/Z=z)p(z)$

REDUCCIÓN VARIANZA: $Var_Z[E(X/Z)] = Var[X] - E_Z[Var(X/Z)] \leq Var[X]$

PROCEDIMIENTO: MUESTREAR Z Y PARA CADA UNO CALCULAR ANALÍTICAMENTE ESPERANZA CONDICIONADA DE X. PARA $\{z_1,\ldots,z_n\}$ SE CALCULA $\sum_{i=1}^{n} \frac{1}{n} E[X/Z=z_i]$.

EJ.: RESPUESTA INCENDIOS SERIOS (1 DE 30), 30 VECES MÁS DE NECESARIAS. POSIBLE SABER TIEMPO RESPUESTA POR POSICIÓN VEHÍCULOS: SIMULAR COMO SI HUBIERA.

PROBLEMA: Z CON ESPERANZA CONDICIONADA ANALÍTICA? NO SIEMPRE POSIBLE

6.5 MUESTREO ESTRATIFICADO

OBJETIVO: ELIMINAR PARTE DE LA VARIABILIDAD MUESTREANDO EN LOS DISTINTOS ESTRATOS O ESCENARIOS POSIBLES

IDEA BÁSICA: h(X): VARIABLE DE SALIDA E[h (X)]= μ**? X V.A. DENSIDAD** $f(x),\, x \in D$.

PARTICIÓN SOPORTE ESTRATOS: $D = \bigcup_i D_i, \quad D_i \cap D_j = \varnothing$ **ESPERANZA ESTRATO:**

$$\mu_i = \int_{D_i} h(x) f(x) dx$$

ESTIMADOR MUESTREO ESTRATIFICADO: $\mu = \int_D h(x) f(x) dx = \sum_{i=1}^{k} \int_{D_i} h(x) f(x) dx = \sum_{i=1}^{k} \mu_i$

MENOR VARIANZA POR MENOR VARIABILIDAD EN LOS ESTRATOS.

PROCEDIMIENTO:

- DIVIDIR SOPORTE EN ESTRATOS (ESCENARIOS)

- ESTIMAR EL VALOR EN CADA ESTRATO

- SUMARLOS (CONCENTRAR PUNTOS MUESTRALES ESTRATOS MÁS IMPORTANTES)

PROB.: ELEGIR ESTRATOS, TAMAÑO MUESTRA ESTRATO. GENERAR MUESTRAS ESTRATO

6.6 MUESTREO POR IMPORTANCIA

OBJETIVO: MUESTREAR PUNTOS MAYOR IMPORTANCIA O INFLUENCIA EN ESTIMACIÓN

IDEA BÁSICA: X V.A. : $f(x)$**,** $x \in D \subset \mathbb{R}^m$ $\longrightarrow$ **ESTIMAR** $\mu = E[h(X)] = \int_D h(x)f(x)dx$

Z V.A. F. DENSIDAD g **DOMINIO** D **(DIST. IMPORTANCIA)** $\mu = \int_D \frac{h(z)f(z)}{g(z)} g(z)dz = E_Z\left(\frac{h(z)f(z)}{g(z)}\right)$

ESTIMADOR POR IMPORTANCIA (MUESTREAR EN z_i **):** $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(z_i)f(z_i)}{g(z_i)}$

(MEDIA DE LOS $h(z_i)$ **PONDERADOS CON PESOS** f/g**)**

ÚTIL SI DISTRIBUCIÓN DE IMPORTANCIA CON FORMA SIMILAR A hf**.**

CONCENTRAR DISTRIBUCIÓN EN PUNTOS DE MAYOR IMPORTANCIA

(SI SUPIÉSEMOS DE ANTEMANO QUE ALGUNOS VALORES SON MÁS IMPORTANTES QUE OTROS EN DETERMINACIÓN DE PARÁMETRO, SELECCIONAR ÉSTOS CON MÁS FRECUENCIA)

PROBLEMA: DIFÍCIL SABER PUNTOS IMPORTANTES, MUY DEPENDIENTE DE MODELO (PRUEBA PILOTO)

6. DISEÑO DE EXPERIMENTOS

COMPARAR DIFERENTES ALTERNATIVAS:

FACTORES CONTROLABLES OPERAN A DISTINTOS NIVELES Y OTROS NO CONTROLABLES

INTERESA IDENTIFICAR EL EFECTO DE CADA FACTOR, LA INTERACCIÓN ENTRE FACTORES, ETC.

¿QUÉ PRUEBAS O REPLICACIONES A HACER PARA IDENTIFICARLO? $\rightarrow$

DISEÑO DE EXPERIMENTOS

7.1. PRINCIPIOS DEL DISEÑO EXPERIMENTAL
7.2. DISEÑO CON UN FACTOR
7.3. DISEÑO EN BLOQUES ALEATORIZADOS
7.4. MODELOS FACTORIALES: DOS FACTORES E INTERACCIÓN
7.5. MODELOS FACTORIALES CON MÁS DE DOS FACTORES
7.6. CUADRADOS LATINOS
7.7. MODELOS CON EFECTOS ALEATORIOS: DISEÑOS JERÁRQUICOS
7.8. DISEÑOS FACTORIALES A DOS NIVELES (2^k)
7.9. MODELO DE REGRESIÓN O SUPERFICIE DE RESPUESTA

7.1. PRINCIPIOS DEL DISEÑO EXPERIMENTAL

OBJETIVO EXPERIMENTO: ESTUDIAR EFECTO SOBRE RESPUESTA → VARIABLES EXPERIMENTALES, FACTORES O TRATAMIENTOS.

HIPÓTESIS: V. RESPUESTA CONTINUA, FACTORES FIJADOS NIVELES

NÚMERO TOTAL DATOS: TAMAÑO DEL EXPERIMENTO.

FORMAS DE ELIMINAR EFECTO OTROS FACTORES NO INTERÉS:

1) MANTENER FIJO EL NIVEL DURANTE TODO EL EXPERIMENTO

2) REORGANIZAR ESTRUCTURA EXPERIMENTO LAS COMPARACIONES DE INTERÉS SE EFECTÚEN PARA VALORES FIJOS DE ESTA VARIABLE

3) ALEATORIZAR SU APARICIÓN EN LOS TRATAMIENTOS

(1 Y 2 CONTROLABLES, 3 NO CONTROLABLE Y POCA INFLUENCIA: ERROR)

EL PRINCIPIO DE ALEATORIZACIÓN

FACTORES NO CONTROLADOS ASIGNAR AL AZAR A OBSERVACIONES
(V. RESPUESTA: VENTAS. FACTOR: DÍA. FACTOR NO CONT: EMPLEADO)

CONTRASTES DE PERMUTACIONES PARA FACTOR EN PRESENCIA DE OTRA VARIABLE

EJ: OBSERVAR 5 DÍAS UN TIPO (D1) Y 5 DE OTRO (D2).

1. ASIGNAR ALEATORIAMENTE 2 EMPLEADOS A CADA DÍA (5 DÍAS CADA UNO)

2.OBSERVAR DIFERENCIA MEDIAS VENTAS: $\overline{x}_{D_1} - \overline{x}_{D_2}$.

3. SI DÍA NO INFLUYE, DA IGUAL OBSERVACIONES UN DÍA U OTRO

POSIBLES ASIGNACIONES DÍA A 10 OBSERVACIONES: $\binom{10}{5} = 252$

CALCULAR 252 DIFERENCIAS, Y SITUAR VALOR ANTERIOR (¿RARO?)

LA REPETICIÓN DEL EXPERIMENTO

REPETIR CADA OBSERVACIÓN $\neq$ REPETIR EL EXPERIMENTO.

DOS MEDIDAS SEGUIDAS: OBSERVACIÓN REPETIDA DOS VECES

REALIZAR EXPERIMENTO UNA VEZ Y REPETIRLO DESDE EL PRINCIPIO

HOMOGENEIDAD ESTAD. COMPARACIONES: DISEÑOS FACTORIALES

ENFOQUES PARA LOGRAR HOMOGENEIDAD:

1. DISEÑOS CLÁSICOS: ELIMINAR LAS DEMÁS VARIABLES (FIJAS A NIVELES CONSTANTES Y VARIAR SÓLO FACTOR)
2. DISEÑOS FACTORIALES: INTRODUCIR TODAS LAS VARIABLES Y ESTIMAR EFECTOS PROMEDIANDO SITUACIONES HOMOGÉNEAS.

INCONVENIENTE CLÁSICO: NO TIENE EN CUENTA INTERACCIÓN

EL CONCEPTO DE BLOQUE

VARIABLE BLOQUE: VARIABLE CUYO EFECTO NO ES DIRECTAMENTE DE INTERÉS UTILIZADA PARA OBTENER COMPARACIONES HOMOGÉNEAS

SE SUELE SUPONER INDEPENDENCIA FACTORES Y VARIABLES BLOQUE.

7.2. DISEÑO CON UN FACTOR

UN FACTOR QUE OPERA A I NIVELES:

TAMAÑO: n **(**n_i **CON FACTOR AL NIVEL** i**,** y_{ij} **OBSERVACIONES)**

MODELO: $y_{ij} = \mu_i + u_{ij}$, $\quad u_{ij} \overset{D}{=} N(0,\sigma)$ **INDEPENDIENTES**

PARÁMETROS: μ_i **Y** σ^2

1) ESTIMAR PARÁMETROS

A) DEL FACTOR $\hat{\mu}_i = \overline{y}_{i.} = \dfrac{\sum_j y_{ij}}{n_i}$

B) (RESIDUOS $e_{ij} = y_{ij} - \overline{y}_{i.}$**), DE LA VARIANZA** $\hat{\sigma}^2 = \dfrac{\sum_{ij} e_{ij}^2}{n}$**.**

2) CONTRASTE IGUALDAD DE MEDIAS: $\mu_1 = \ldots = \mu_I = \mu$**:**

$$H_0 : \mu_1 = \ldots = \mu_I = \mu$$

ANÁLISIS DE LA VARIANZA: COMPARA VARIABILIDAD EXPLICADA POR GRUPOS (VE) CON NO EXPLICADA O RESIDUAL (VNE) (SI MUY GRANDE RECHAZAR H0)

Fuentes variación	Suma de cuadrados	Grados de libertad	Varianzas
Entre grupos (VE)	$\sum_i n_i(\overline{y}_{i.} - \overline{y}_{..})^2$	$I-1$	$\hat{s}_e^2$
Interna, no explicada o residual (VNE)	$\sum_{i,j} (y_{ij} - \overline{y}_{i.})^2$	$n-I$	$\hat{s}_R^2$
Total	$\sum_{i,j} (y_{ij} - \overline{y}_{..})^2$	$n-1$	$\hat{s}_y^2$

$$F_{(I-1),(n-I)} = \hat{s}_e^2 \,/\, \hat{s}_R^2$$

3) COMPROBAR CON RESIDUOS HIPÓTESIS DEL MODELO

4) SI SE HA RECHAZADO H0, COMPARACIONES DE LAS MEDIAS (ENTRE GRUPOS): INTERVALOS CONFIANZA DIFERENCIAS 2 A 2.

MÉTODO BONFERRONI: NIVEL SIGNIFICACIÓN α / c (c Nº COMPARACIONES). EVITAR SÍ DIFERENCIAS SI NO LAS HAY.

7.3. DISEÑO EN BLOQUES ALEATORIZADOS

ANALIZAR EFECTO DE UN FACTOR PRESENCIA VARIABLE BLOQUE:

I **NIVELES FACTOR,** J **VALORES VAR. BLOQUE,** $I \cdot J$ **OBSERV.** y_{ij}

$y_{ij} = \mu + \alpha_i + \beta_j + u_{ij}$**,** $u_{ij}\ N(0,\sigma)$ **INDEPTES. (**$\sum_i \alpha_i = 0$ **,** $\sum_j \beta_j = 0$**)**

1) ESTIMAR PARÁMETROS

$$\hat{\alpha}_i = \overline{y}_{i.} - \overline{y}_{..} \qquad \hat{\beta}_j = \overline{y}_{.j} - \overline{y}_{..} \qquad e_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i,j} e_{ij}^2$$

2) CONTRASTES (ANÁLISIS DE LA VARIANZA)

A) $H_0 : \alpha_i = 0; \quad \forall i$ **B)** $H_0 : \beta_j = 0; \quad \forall j$

3) COMPROBAR CON RESIDUOS HIPÓTESIS DEL MODELO

INTERACCIÓN: $y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ij}$ **(RESIDUOS NO MEDIA 0. GRÁF: X: VALOR PREVISTO Y: RESIDUO, CURVATURA)**

4) SI SE HA RECHAZADO H0, COMPARACIONES DE LAS MEDIAS

7.4. MODELOS FACTORIALES: DOS FACTORES E INTERACCIÓN

MODELO SIN REPLICACIÓN: $y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ij}$

($\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$).

$IJ + 1$ **PARÁMETROS (NO SUF. OBSERVACIONES)** $\rightarrow$

- **A) NO ESTIMAR VARIANZA (NO ANOVA)**
- **B) SUPONER NO INTERACCIÓN**
- **C) REPLICAR EL MODELO:**

MODELO C/REPLICACIÓN: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ijk}$ (IJK **OBS.**)

7.5. MODELOS FACTORIALES CON MÁS DE DOS FACTORES

MODELO 3 TRES FACTORES:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + u_{ijk}$$

- EN GENERAL: EF. PRINCIPALES > SEGUNDO ORDEN > TERCER ORDEN
- MÁS PARÁMETROS QUE OBSERVACIONES (IGUAL QUE EN 2 FACT.)
- M FACTORES: INTERACCIONES MÁS DE TERCER ORDEN NULAS

PROBLEMA: GRAN NÚMERO DE OBSERVACIONES REQUERIDAS

SOLUCIÓN: NO USAR DISEÑOS COMPLETOS $\rightarrow$ FRACCIÓN:

INTERACCIONES NULAS A PARTIR DE UN ORDEN, ELEGIR UNA PARTE DEL DISEÑO PERMITA ESTIMAR EF. PRINCIPALES E INTERACCIONES

7.6. CUADRADOS LATINOS

A) TRES FACTORES (ALGUNO PUEDE SER VARIABLE BLOQUE);

B) MISMO NÚMERO DE NIVELES PARA CADA FACTOR

C) NO SE ESPERA INTERACCIÓN ENTRE LOS FACTORES

N: Nº NIVELES; DIS. COMPLETO N^3 OBS; CUADRADO LATINO N^2

IDEA BÁSICA: SELECCIÓN CONJUNTO COMBINACIONES DE NIVELES CADA NIVEL DE UN FACTOR APARECE UNA VEZ CON CADA NIVEL OTROS.

PROCEDIMIENTO:

1) ASIGNAR UN FACTOR A FILAS, OTRO COLUMNAS Y OTRO LETRAS
2) SELECCIONAR CUADRADO LATINO DEL NÚMERO DE NIVELES
3) ALEATORIZAR ORDEN FILAS Y COLUMNAS $\rightarrow$ COMB. A EXPERIM.

POSIBLES CUADRADOS LATINOS CON 3 Y CON 4 NIVELES:

$$N=3 \quad \begin{array}{ccc} A & B & C \\ B & C & A \\ C & A & B \end{array} \qquad \begin{array}{ccc} A & B & C \\ C & A & B \\ B & C & A \end{array}$$

$$N=4 \quad \begin{array}{cccc} A & B & C & D \\ B & A & D & C \\ C & D & A & B \\ D & C & B & A \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ D & C & B & A \\ B & A & D & C \\ C & D & A & B \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ C & D & A & B \\ D & C & B & A \\ B & A & D & C \end{array}$$

EJ: EN CONSUMO AUTOM. 4 TIPOS GAS., 4 AUTOM. Y 4 CONDUCTORES. $N=4$

DISEÑO CUADRADO LATINO:

COLUMNAS: CONDUCTORES. FILAS: VEHÍCULO. LETRAS: TRATAMIENTOS GAS.

1) CUADRADO LATINO SELECCIONADO:

$$\begin{array}{cccc} A & B & C & D \\ B & A & D & C \\ C & D & A & B \\ D & C & B & A \end{array} = \begin{array}{cccc} T1 & T2 & T3 & T4 \\ T2 & T1 & T4 & T3 \\ T3 & T4 & T1 & T2 \\ T4 & T3 & T2 & T1 \end{array}$$

2) ALEATORIZAR: F (2,1,3,4), C (3,4,2,1)

	C3	C4	C2	C1
V2	T1	T2	T3	T4
V1	T2	T1	T4	T3
V3	T3	T4	T1	T2
V4	T4	T3	T2	T1

REORDENANDO

	C1	C2	C3	C4
V1	T3	T4	T2	T1
V2	T4	T3	T1	T2
V3	T2	T1	T3	T4
V4	T1	T2	T4	T3

3) MODELO: $y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_k + u_{ij(k)}, \quad i, j, k = 1, \ldots, N$

α_i**: EFECTO FILA,** β_j **EFECTO COLUMNA,** γ_k **EFECTO LETRA,** $u_{ij(k)}$ $N(0, \sigma^2)$ **IND.**

($ij(k)$ **INDICA QUE** k **DEPENDE DE CASILLA** ij **SEGÚN DISEÑO)**

$$\sum_1^N \alpha_i = \sum_1^N \beta_j = \sum_1^N \gamma_k = 0.$$

PARA CUATRO FACTORES: CUADRADOS GRECO-LATINOS

7.7. MODELOS CON EFECTOS ALEATORIOS: DISEÑOS JERÁRQUICOS

DISEÑO EFECTOS FIJOS: EFECTO FACT. A DETERMINADOS NIVELES

DISEÑO EFECTOS ALEATORIOS: LOS NIVELES NO SE ELIGEN

MEDIR VARIABILIDAD RESPUESTA (NO VARIACIÓN NIVEL)

EJ.: SABER SI CIERTOS TIPOS CARBURANTE EFECTIVOS: EF. FIJOS

SABER VARIABILIDAD POR USO DISTINTOS CARBURANTES: EF. ALEAT.

OBJETIVO: ESTIMAR VARIANZA FACTOR, NO α SINO σ_α^2

- MODELO: IGUAL, EF. FIJOS RESPUESTA MEDIA PARÁM. A ESTIMAR,

EF. ALEATORIOS V.A. N(0, σ_α^2), A ESTIMAR.

- ANOVA IGUAL. CONTRASTES SIN INTERACCIÓN IGUAL.

CON INTERACCIÓN, EF. FIJOS CONTRASTA $\hat{s}_\alpha^2 / \hat{s}_R^2$, EF. ALEAT. $\hat{s}_\alpha^2 / \hat{s}_{\alpha\beta}^2$.

- NO PROBLEMA COMPARACIONES MÚLTIPLES

CASO ESPECIAL: FACTORES ANIDADOS O JERARQUIZADOS:

EJ: INFLUENCIA EN OPINIÓN TRABAJ. DE SECTOR INDUSTRIAL Y DE EMPRESA

1) SECTORES 2) POR SECTOR SELEC. EMPRESAS 3) POR EMPRESA SELEC. TRABAJ.

FACTOR EMPRESA PUEDE MEDIRSE SÓLO A UN NIVEL DEL SECTOR.

PRINCIPAL DISEÑO ANIDADO: MODELO COMPONENTES DE VARIANZA

VARIABILIDAD EN RENDIMIENTO DE MÁQUINAS Y OPERARIOS. SELEC. MÁQUINAS AZAR Y MUESTREAR TRABAJADORES. NO TODAS LAS MÁQUINAS Y DIST. TRABAJ.

MODELO: $y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + u_{ijk}$, α_i **EF. ALEAT. MÁQUINA** $N(0,\sigma_\alpha^2)$, $\beta_{j(i)}$ **EF. TRABAJADOR** j **CON MÁQUINA** i $N(0,\sigma_\beta)$, u_{ijk} **ERROR** $N(0,\sigma)$

$\sigma_y^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$. **ESTIMACIÓN SECUENCIAL**

7.8. DISEÑOS FACTORIALES A DOS NIVELES (2^k)

DISEÑOS FACTORIALES A DOS NIVELES: GRAN NÚMERO FACTORES FÁCILES FRACCIONAR

1. **DISEÑOS FACTORIALES COMPLETOS**
2. **CÓMO FRACCIONARLOS Y APROVECHAR PARA EXP. SECUENCIAL.**

DISEÑO FACTORIAL 2^k: k FACTORES 2 NIVELES CADA UNO.

EL DISEÑO 2^2

FACTORES: A Y B. NOTACIÓN NIVELES: + Y –

TABLA OBSERVACIONES: (O) - Y – , (A) A + B –, (B) A - B +, (AB) + Y +

A	B	Y
–	–	y_{11} (o)
+	–	y_{21} (a)
–	+	y_{12} (b)
+	+	y_{22} (ab)

MODELO: $X_k = \begin{cases} +1 & \text{si factor } k \text{ en nivel } + \\ -1 & \text{si factor } k \text{ en nivel } - \end{cases},\quad y_{ij} = \mu + \alpha_2 X_1 + \beta_2 X_2 + (\alpha\beta)_{22} X_1 X_2 + u_{ij}; \quad i,j = 1,2$

$$\alpha_2 = -\alpha_1 = \frac{1}{2}\alpha \qquad \beta_2 = -\beta_1 = \frac{1}{2}\beta \qquad (\alpha\beta)_{22} = \frac{1}{2}(\alpha\beta)$$

ESTIMACIÓN: $\hat{\mu} = \frac{1}{4}(o+a+b+ab)$ $\hat{\alpha} = \frac{1}{2}(a+ab-o-b)$ $\hat{\beta} = \frac{1}{2}(b+ab-o-a)$ $\widehat{(\alpha\beta)} = \frac{1}{2}(o+ab-a-b)$

ALGORITMO SIGNOS:

1) TABLA ESTÁNDAR: AÑADIR COLUMNA AB PRODUCTO DE A Y B

2) ESTIMACIÓN: MULTIPLICAR COLUMNA SIGNOS POR OBS. Y /2.

A	B	AB	Y
–	–	+	y_{11} (o)
+	–	–	y_{21} (a)
–	+	–	y_{12} (b)
+	+	+	y_{22} (ab)

4 PARÁMETROS 4 OBSERVACIONES: NO ESTIMACIÓN ERROR

SI INTERACCIÓN = 1/3(PROMEDIO EF. PRINC.) INADECUADA.

VENTAJAS: ESTIMACIONES MÁS PRECISAS (MENOR VARIANZA)

EL MODELO 2^3

3 FACT. 2 NIVELES: 3 EF. PRINC., 3 INTER. 2º ORDEN, 1 INTER.3º

A	B	C	AB	AC	BC	ABC	Y
–	–	–	+	+	+	–	(o)
+	–	–	–	–	+	+	(a)
–	+	–	–	+	–	+	(b)
+	+	–	+	–	–	–	(ab)
–	–	+	+	–	–	+	(c)
+	–	+	–	+	–	–	(ac)
–	+	+	–	–	+	–	(bc)
+	+	+	+	+	+	+	(abc)

ESTIMACIÓN: SIGNOS Y DIVIDIR $n/2 = 2^{k-1} = 2^2$**. (MEDIA** $n = 2^k$**)**

ESTIMAR VARIANZA RESIDUAL:

A) ESTIMACIÓN EXTERNA

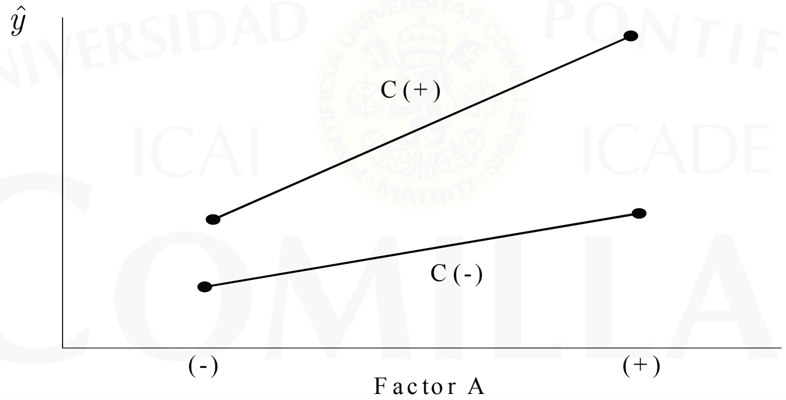
B) REPLICAR DISEÑO

C) CONSIDERAR NULA INTERACCIONES: ESTIM. SERÍAN N(0, $\sigma^2/2$**), SI ALGUNA** $\neq$ **0,**

CAMBIA MEDIA NO VARIANZA. MEDIANA ESTIM. INTERAC. $M(\hat{\vartheta}) = mediana(\hat{\vartheta}_5, \hat{\vartheta}_6, \hat{\vartheta}_7, \hat{\vartheta}_8)$

ESTIM. ROBUSTO VARIABIL.: $MEDA = \underset{i=5,\ldots,8}{mediana}\left|\hat{\vartheta}_i - M(\hat{\vartheta})\right|$**. ESTIM. ROBUSTO DESV. TÍP. NORMALES** $\hat{s}_{\vartheta} = \dfrac{MEDA}{0.675}$**.**

CONTRASTES: EF. PRINC. EXISTE SI $\left|\hat{\vartheta}_j\right| \geq 2\hat{s}_{\vartheta}$**, EST. ERROR** $\hat{\sigma} = \hat{s}_{\vartheta}\sqrt{2}$**.**



¿INTERACCIÓN NULA?

A) UNA RECTA SUPERIOR A OTRA: C INFLUYE MUCHO

B) PUNTOS DE DERECHA SIEMPRE SUPERIORES: A TAMBIÉN INFLUYE

C) NO PARALELAS: INTERACCIÓN

EL MODELO 2^k

PRÁCTICAMENTE IGUAL QUE EL MODELO ANTERIOR.

MATIZACIONES:

1) INTERACCIONES ALTAS SE SUELEN CONSIDERAR NULAS (>3)

ESTIMADOR A PARTIR DE MEDA PARA ERROR EXPERIMENTAL:

$\hat{\sigma} = \hat{s}_{\vartheta}\sqrt{2^{k-2}}$.

CONTRASTES: SIGNIFIC. SI $\left|\hat{\vartheta}_j\right| \geq 2\hat{s}_{\vartheta}$ **(SI** k **GRANDE CON 3)**

2) ¿MEJOR DISEÑO 2^{k-1} **REPLICADO O INTRODUCIR FACTOR** 2^k**?**

(RIESGO DE AÑADIR FACTOR INERTE)

MEJOR 2^k**: SI INERTE PROCEDIMIENTOS IGUALES (REPLICADO), SI ACTIVO, PERMITIRÁ MEDIR SU EFECTO.**

FRACCIONES DE DISEÑOS FACTORIALES A DOS NIVELES

2^k: MUCHAS PRUEBAS

DIS. FRACCIONALES: UNA FRACCIÓN PARA ESTIMAR PARÁMETROS ORDEN BAJO.

NULAS INTERACCIONES ORDEN ALTO

USO: IDENTIFICAR PRELIMINAR FACT.

EJ: 2^3 FRAC. ABC +:

A	B	C	AB	AC	BC	ABC	Y
+	–	–	–	–	+	+	(a)
–	+	–	–	+	–	+	(b)
–	–	+	+	–	–	+	(c)
+	+	+	+	+	+	+	(abc)

(1/2 OBSERVACIONES)

ESTIMACIÓN: IGUAL (DIVIDIR POR 2 NO POR 4) PERO:

NO SÓLO EFECTO FACTOR, TAMBIÉN INTERACCIONES: EFECTOS CONFUNDIDOS.

EJ: ESTIM. PAR. DE A INCLUYE INTERACCIÓN BC (MISMOS SIGNOS)

$$\hat{s}_A \to A + BC \quad \hat{s}_B \to B + AC \quad \hat{s}_C \to C + AB \quad \hat{s}_\mu \to \mu + ABC$$

SI FRACCIÓN (-) ABC, EFECTOS CONFUNDIDOS:

$\hat{s}_A \to A - BC \quad \hat{s}_B \to B - AC \quad \hat{s}_C \to C - AB \quad \hat{s}_\mu \to \mu - ABC$

SÓLO ESTIMAR EFECTOS PRINCIPALES SI NULAS INTERACCIONES.

ECUACIÓN GENERATRIZ: PARA SABER QUÉ EFECTOS CONFUNDIDOS

EJ. 1: ECUACIÓN GENERATRIZ $I = ABC$ **(**I **COL. DE +)**

OBTENER EF. CONFUNDIDOS: MULTIPLICAR EC. GEN. POR FACTORES

(REGLAS: $AI = A$ **Y** $AA = I$**)**

EJ.: $A = AI = AABC = IBC = BC$

$B = BI = BABC = AC$ **Y** $C = IC = ABCC = ABI = AB$

COMPARAR FRACCIONES:

RESOLUCIÓN = 1+ORDEN INTER. MÁS BAJA CONFUNDIDA EF. PRINC.

EJ: MEDIA FRACCIÓN DE 2^5 **(**2^{5-1}**) CON** $I = ABCDE$ **RESOLUCIÓN 5**

(EF. PRINC. CONF. INTER. 4º ORDEN); CON $I = ABCD$ **4.**

RESOLUCIÓN = NÚMERO DE LETRAS EN ECUACIÓN GENERATRIZ EN PAL. MÁS CORTA

CONSTRUCCIÓN DE DISEÑOS 2^{k-p} EJ: 5 FACT. 8 EXP. $\rightarrow$ 2^{5-2}.

PROCEDIMIENTOS:

1) ELEGIR DE 2^5 QUÉ DOS INTERACCIONES PONER TODAS + O –

2) HACER UN 2^3 Y ESTUDIAR CÓMO AÑADIR NUEVOS 2 FACT.: MEJOR PUES EN 1) DIFÍCIL DETERMINAR RESOLUCIÓN: $I = ABCDE$, $I = ABCD$ $\longrightarrow$ $E = EI = EABCD = ABCDE = I$ RES=1

¿CÓMO AÑADIR VARIABLES?: ASIGNAR VARIABLE A UNA INTER. ALTA.

EJ.: $D = ABC$ Y $E = AB$, ECUAC. GENER. $I = ABCD = ABE$, RES=3

NO POSIBLE NO ESTIMAR EF. PRINC. UN FACTOR, Y DIRECTAMENTE EC. GENERATRIZ

EC. GEN. COMPLETA: TODAS POSIBLES MULTIPLICACIONES DE EC. DE SELECCIÓN

EJ.: EN 2^{6-3} CON $D = AB$ $\quad E = AC$ $\quad F = BC$, EC. GEN.: $I = ABD = ACE = BCF$

" COMPLETA: $I = ABD = ACE = BCF = BDCE = ACDF = ABEF = DEF$.

ECUACIÓN GENERATRIZ COMPLETA $\rightarrow$ CONFUSIÓN MULT. POR EFECTO

EJ: EF. CONF. CON A: $A = BD = CE = ABCF = ABDCE = CDF = BEF = ADEF$

ANÁLISIS DE FRACCIONES FACTORIALES

$\hat{s}_{\vartheta} = MEDA / 0.675$ SIGNIFICATIVO SI $> 2\hat{s}_{\vartheta}$ ($3\hat{s}_{\vartheta}$ MUCHOS)

DECIDIDAS SIGNIF., DECIDIR EFECTOS CONFUNDIDOS EN ELLAS SIGNIFICATIVO:

EF. PRINC. > INTER.; INTER. ORDEN L > ORDEN L+1; RARO ALTAS ACT.

EJ: 2^{3-1}, $I = ABC$ $\longrightarrow$ $\hat{s}_A \rightarrow A+BC$ $\hat{s}_B \rightarrow B+AC$ $\hat{s}_C \rightarrow C+AB$ $\hat{s}_\mu \rightarrow \mu+ABC$

SI SÓLO ACTIVA $\hat{s}_A$: A) SÓLO A ACTIVA

B) SÓLO BC ES ACTIVA

C) AMBAS ACTIVAS

APLICACIONES DISEÑOS FACTORIALES

1) EXP. OFF-LINE: PERSONAL ESPEC., GRAN Nº VAR., TIEMPO LIM. $\longrightarrow$ DISEÑOS FRACCIONALES

2) EXP. ON-LINE: PERSONAL NO ESPEC., SENCILLA, NO LIM. TIEMPO $\longrightarrow$ DISEÑOS SIMPLES (2^2 O 2^3) REPLICADOS

EXPERIMENTACIÓN ON-LINE: OPERACIÓN EVOLUTIVA O EVOP

EJ. 2 FACTORES:

1) SELEC. 2 NIVELES CADA Y EXPERIMENTAR CON 2^2

2) SI PASADO TIEMPO SE DECIDE FACTOR A SIGNIF.:

   SITUAR A EN MEJOR NIVEL.

   SELECCIONAR OTRO NIVEL PARA A

   REPETIR HASTA CONSIDERAR NO MEJORABLE CONFIGUR.

7.9. MODELO DE REGRESIÓN O SUPERFICIE DE RESPUESTA

FACTORES CONTINUOS: EFECTO IMPORTANTES, RESTANTES ERROR

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \beta_{jh} x_{ij} x_{ih} + \varepsilon_i, \quad \varepsilon_i \text{ v.a.i.i.d. } N(0, \sigma^2)$$

¿VALORES? $\rightarrow$ ESTIMACIÓN

¿SIGNIFICATIVOS? $\rightarrow$ ANÁLISIS DE LA VARIANZA.

¿VÁLIDO? $\rightarrow R^2$ Y ANÁLISIS RESIDUOS (NORMALIDAD, HOMOCEDASTICIDAD,...)

PROCEDIMIENTOS ESTIMAR MODELO:

1) ESTIMAR PARÁMETROS CONJUNTAMENTE SIGNIFIC. Y VOLVER ESTIMAR SIN NO SIGNIF (EFECTOS CONJUNTOS REPARTIDOS)

2) REGRESIÓN POR PASOS: AÑADIR O QUITAR 1 VARIABLE HASTA QUE NO INCLUIDAS NO APORTAN INFORMACIÓN (CORRELACIÓN CON RESIDUOS,COEF. NO INCLUYE INTER.)

TAMAÑO EXPERIMENTACIÓN: NO RECOMENDABLE Nº VAR. TAL QUE Nº VAR./Nº DATOS ALTO $\rightarrow R^2$ TIENDE A SER ALTO.

CONSTRUCCIÓN DE MODELOS DE SIMULACIÓN VÁLIDOS Y CREÍBLES

EL OBJETIVO PRINCIPAL DE LA METODOLOGÍA DE MODELADO ES QUE UN MODELO DEBE SER UNA REPRESENTACIÓN ADECUADA DEL SISTEMA QUE SE ESTUDIA.

PROBLEMA: ¿CUÁNDO UN MODELO DE SIMULACIÓN ES UNA REPRESENTACIÓN SUFICIENTEMENTE BUENA DEL SISTEMA, I.E., CUÁNDO EL MODELO ES VÁLIDO?

NO HAY RESPUESTA EXACTA A LA PREGUNTA. SÓLO SE PUEDEN DAR ALGUNAS PAUTAS.

1. DEFINICIONES

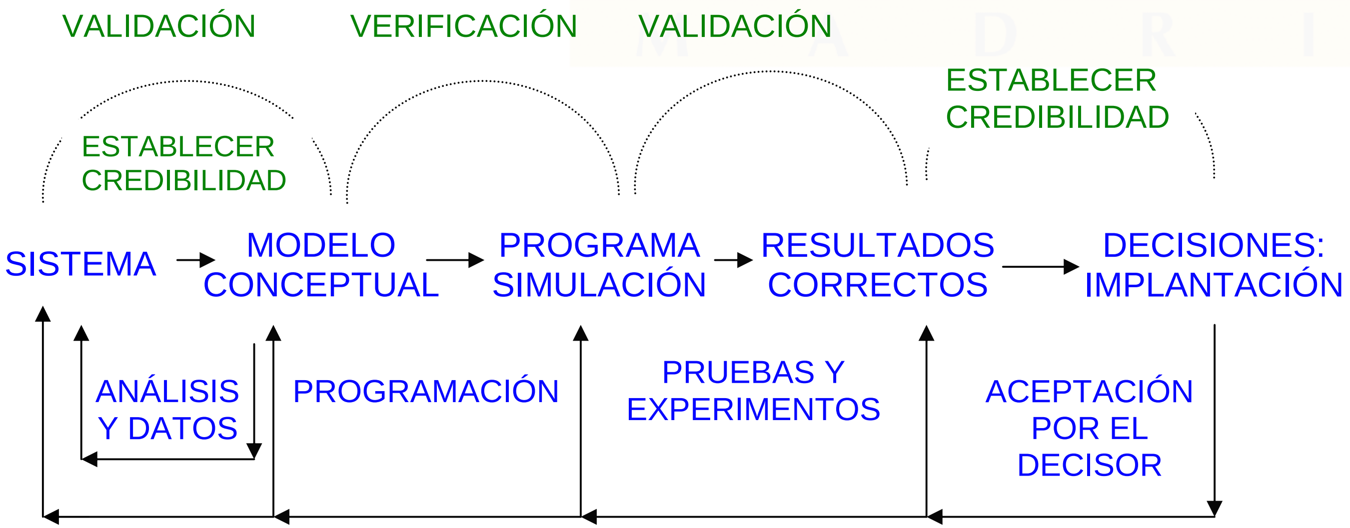
FASES:

- FASE DE ANÁLISIS Y MODELADO → MODELO CONCEPTUAL: REPRESENTACIÓN MATEMÁTICA O LÓGICA DEL PROBLEMA, FORMULADO PARA UN ESTUDIO PARTICULAR.
- FASE PROGRAMACIÓN E IMPLANTACIÓN → MODELO COMPUTACIONAL: TRADUCCIÓN DE MODELO CONCEPTUAL A PROGRAMA DE ORDENADOR E IMPLANTACIÓN EN ORDENADOR
- FASE DE EXPERIMENTACIÓN → INFERENCIAS SOBRE EL PROBLEMA MEDIANTE EXPERIMENTOS COMPUTACIONALES.

VERIFICACIÓN: *DETERMINAR QUE EL PROGRAMA DE ORDENADOR SE COMPORTA COMO ES DEBIDO, I.E., QUE SE HA REALIZADO UNA TRADUCCIÓN CORRECTA DEL MODELO CONCEPTUAL A UN PROGRAMA DE ORDENADOR QUE FUNCIONA CORRECTAMENTE.*

VALIDACIÓN: *DETERMINAR SI EL MODELO CONCEPTUAL ES UNA REPRESENTACIÓN ADECUADA DEL SISTEMA, I.E., SI NUESTRO CONOCIMIENTO DEL SISTEMA SE HA TRADUCIDO EN HIPÓTESIS QUE REPRODUCEN CORRECTAMENTE SU COMPORTAMIENTO REFERENTE A LOS OBJETIVOS DEL ESTUDIO.*

CREDIBILIDAD: *LOS RESULTADOS HAN DE SER ACEPTADOS POR EL DECISOR PARA SER UTILIZADOS EN EL PROCESO DE TOMA DE DECISIONES, I.E., HAN DE SER CREÍBLES.*



- **VALIDACIÓN:**
- **PRIMERO→ DETERMINAR SI HEMOS INCORPORADO TODOS LOS ASPECTOS DEL SISTEMA QUE SON DE REAL INTERÉS PARA EL ESTUDIO Y SÓLO LOS RELEVANTES: EXPERIMENTAR CON EL MODELO ES COMO HACERLO CON UN SUSTITUTO DEL SISTEMA**
- **MODELOS CON OBJETIVO →CLARIDAD OBJETIVOS, IMPLICAR DECISOR (CREÍBLE)**

  EQUILIBRIO ENTRE REALISMO (COMPLEJIDAD) Y OBJETIVOS, NO INCLUIR DETALLES INNECESARIOS →APROXIMACIÓN.

  PROCEDIMIENTOS JERÁRQUICOS: 1º SENCILLO→IR COMPLICANDO.
- **VERIFICACIÓN:**

1) Organizar el programa en submódulos, verificables por separado.
2) **LIBRE DE ERRORES DE PROGRAMACIÓN PROBARLO CON CASOS SENCILLOS, CUYO RESULTADO SEA CONOCIDO.**
3) **TENER UNA TRAZA DE LA EJECUCIÓN PARA SEGUIR LA LÓGICA DE LOS PROCESOS.**
4) **ANIMACIÓN GRÁFICA PARA VERIFICACIÓN MEDIANTE VISUALIZACIÓN Y CREDIBILIDAD**

2. METODOLOGÍA DE VALIDACIÓN Y CREDIBILIDAD

PRIMER PASO: DESARROLLAR DESDE EL PRINCIPIO UN MODELO VÁLIDO (OBJETIVO DE SER BUENO) → CREIBLE, ES DECIR, RAZONABLE.

- INVOLUCRAR A LOS USUARIOS
- VERIFICAR LA RECOGIDA DE DATOS
- SI ES POSIBLE, INTENTAR BASAR ASPECTOS DEL MODELO EN TEORÍAS BIEN ESTABLECIDAS.

SEGUNDO PASO: VERIFICAR EMPÍRICAMENTE HIPÓTESIS EN QUE SE BASA EL MODELO.

- VERIFICAR ADECUADAMENTE LA BONDAD DE AJUSTE.
- ANÁLISIS DE SENSIBILIDAD: IDENTIFICAR CUÁN SENSIBLES SON LOS RESULTADOS DE LA SIMULACIÓN A ALGÚN ASPECTO DEL MODELO, TENIENDO PUÉS ESPECIAL CUIDADO CON LOS MÁS RELEVANTES → DISEÑO DE EXPERIMENTOS.

TERCER PASO: DETERMINAR CUÁN REPRESENTATIVOS SON RESULTADOS DE SIMULACIÓN

- COMPARACIÓN ENTRE RESULTADOS REALES Y RESULTADOS DE SIMULACIÓN.

COMPARACIÓN ENTRE DATOS REALES Y SIMULADOS

- **INSPECCIÓN:** $\{R_1,\ldots,R_k\}$ **REALES,** $\{\pi_1,\ldots,\pi_k\}$ **SIMULADOS.**

1) **PRIMERA IDEA: TESTS ESTADÍSTICOS DE HOMOGENEIDAD** $\rightarrow$ **NO ES BUENO, PUES NO SUELE HABER INDEPENDENCIA.**

2) **COMPARACIONES DIRECTAS DE MEDIAS, VARIANZAS, ETC.** $\rightarrow$ **PELIGROSO, POR VULNERABILIDAD A LA ALEATORIEDAD.**

- **INSPECCIÓN CORRELACIONADA: ESTABLECER LAS COMPARACIONES ALIMENTANDO EL MODELO CON LOS DATOS HISTÓRICOS.**

DATOS HISTÓRICOS DE ENTRADA AL SISTEMA $\rightarrow$ **SISTEMA REAL** $\rightarrow$ **RESULTADOS SISTEMA**

DATOS HISTÓRICOS DE ENTRADA AL SISTEMA $\rightarrow$ **MODELO DE SIMULACIÓN** $\rightarrow$ **RESULTADOS MODELO**

$\updownarrow$ COMPARAR

- **ANÁLISIS DE INTERVALOS DE CONFIANZA BASADOS EN DATOS INDEPENDIENTES:**

$\{R_1,\dots,R_m\}$ m **CONJUNTOS INDEPENDIENTES DEL SISTEMA**

$\{M_1,\dots,M_n\}$ n **CONJUNTOS INDEPENDIENTES DEL MODELO**

$$\mu_R = E[R_j] \quad \mu_M = E[M_j]$$

INTERVALO DE CONFIANZA PARA: $\xi = \mu_R - \mu_M$

$$\bar{R} - \bar{M} \pm t_{\hat{f},\alpha/2}\sqrt{\frac{S_R^2}{m} + \frac{S_M^2}{n}}; \quad \hat{\mathrm{f}} = \frac{(S_R^2/\mathrm{m} + S_M^2/\mathrm{n})^2}{\dfrac{(S_R^2/\mathrm{m})^2}{m-1} + \dfrac{(S_M^2/\mathrm{n})^2}{n-1}}$$