

Módulo II. Introducción a las técnicas de investigación.

# Introducción a las técnicas de investigación.

# Contenidos

- Tipos de técnicas de investigación.
  - Cualitativos y cuantitativos.
  - Otras clasificaciones o tipos de estudios
    - Revisión sistemática de la literatura
    - Investigación basada en la evidencia
    - Comparativas (*benchmarking*)
- Introducción al diseño de investigaciones.
  - Escalas y medición
  - Tipos de estudios empíricos
    - Encuestas, Casos de estudio, Experimentación formal
  - Proceso y diseño de la experimentación formal

Agradecimientos por parte del material a  
Javier Dolado, Universidad del País Vasco

# Tipos de técnicas de investigación.

# Estudios cualitativos vs. cuantitativos

- Los estudios **cualitativos** buscan la interpretación de un fenómeno en su entorno natural, por ejemplo, recabando información entre las personas involucradas.
- Los estudios **cuantitativos**, por su parte, buscan medir la *influencia* de una variable en un fenómeno, es decir la relación causa-efecto entre ambos (por ejemplo, la relación entre la profundidad en la jerarquía de clases en un diseño orientado a objetos y el número de errores en el código de una clase)

# Investigación basada en la evidencia

- El modelo se apoya en la recogida y análisis sistemáticos de todos los datos empíricos disponibles sobre un determinado fenómeno.
- El **objetivo** último es conocer un fenómeno con mayor profundidad y perspectiva de los que los estudios individuales podrían proporcionar, ya que éstos a menudo están sesgados por factores como el contexto en que se llevan a cabo o las personas que participan.
- Central a la investigación basada en la evidencia es la **revisión sistemática de la literatura**.
  - En medicina:  
[http://es.wikipedia.org/wiki/Medicina\\_basada\\_en\\_la\\_evidencia](http://es.wikipedia.org/wiki/Medicina_basada_en_la_evidencia)
  - En Ingeniería del software:  
<http://www.dur.ac.uk/ebse/>

# Revisión sistemática de la literatura

- Revisiones Sistemáticas (RS) son un diseño de investigación observacional y retrospectivo, que sintetiza los resultados de múltiples investigaciones.
- Los resultados de estudios basados en un único ensayo no pueden considerarse lo suficientemente fiables como para ser generalizados.
- Los estudios individuales que los conforman se denominan *estudios primarios*, por lo que tanto la RS como el **meta-análisis** son considerados *estudios secundarios*.

# Características de la RS

- La RS dispone de un **protocolo** en el que se especifica la pregunta de investigación y los métodos que se utilizarán para llevar a cabo la revisión.
- La RS se basa en una **estrategia de búsqueda** bibliográfica elaborada de forma que se localice toda la literatura relevante posible.
- La estrategia de búsqueda debe estar **definida de forma explícita**, de manera que la RS sea reproducible.
- Deben documentarse los **criterios de inclusión y exclusión** de estudios.
- La RS debe especificar la información obtenida de cada estudio, además de una **evaluación crítica de la calidad** de dichos estudios.
- Debe existir un **proceso transparente de interpretación** de los resultados.

(G. Atienza et al, 2008, Las revisiones sistemáticas, Guías Clínicas  
<http://www.fisterra.com/guias2/FMC/rrss.asp> )

# Comparación (*Benchmarking*)

- La comparación o benchmarking consiste en comparación de unos resultados con otros considerados estándares.
  - En este ámbito, nos referimos más a comparativas de una organización con otros estándares, como la productividad ,p.e.,  
<http://www.isbsg.com/>
  - Puede ir desde la comparativa de un programa con ciertos programas utilizados considerados estándares, p.e., <http://www.phoronix.com/>

# Investigación empírica

- La experimentación provee una manera sistemática, disciplinada, cuantificable y controlada de evaluar actividades desarrolladas por humanos (Wholin et al, 2000).
- Referencias en la ingeniería del software
  - Wohlin C., Runeson P., Höst M., Ohlson M., Regnell B. and Wesslén A. (2000). *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers.
  - Juristo N. y Moreno A.Mª (2001). *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers.
  - Dolado Cosín, J.J. y Fernández Sanz L. (2000). *Medición para la gestión en la Ingeniería del Software*. Ra-Ma.

# Diseño experimental

- Definición de los elementos y las propiedades que deben cumplir los estudios empíricos
- Clasificación los diferentes tipos de diseños experimentales
- Riesgos para la validez interna de cada uno de ellos, con indicación de las técnicas útiles para el control y la medición
- Réplicas de experimentos y familias de estudios

# Escalas y medición

# Teoría de la representación

- La teoría de la representación de la medición establece los principios generales de la medición y su validez.
- Esta teoría trata de expresar de forma numérica (mundo formal) las entidades del mundo real (o mundo empírico) y la correspondencia entre ambos mundos.

# Entidades y atributos

- Se denomina **entidad** a un objeto que va a ser caracterizado mediante una medición de sus atributos. (ISO/IEC 15939)
- Un **atributo** es una característica medible de una entidad.

# Medición

- La **medición** es el proceso por el que se asignan números o símbolos a atributos de entidades del mundo real para describirlos según unas reglas definidas de antemano.
- **Medida** es la asignación de un símbolo o número resultado de una medición a una entidad para caracterizar un atributo.

Fenton, N.E. y Pfleeger S.L., *Software Metrics: A Rigorous and Practical Approach*, 2nd edition, 1998

# Escalas

- Una escala de medición es un conjunto de valores que permite establecer relaciones entre medidas. Con frecuencia dicho conjunto es continuo, está ordenado y viene delimitado por un punto inicial y otro final.

# Tipos de escala

- **Escala nominal.** Formada por categorías entre las cuales no existe ningún orden. P.e., Lenguaje: {Java, C++, Phyton}.
- **Escala ordinal.** Define categorías pero existe una relación de orden "es menor que" entre ellas. P.e., la escala de Likert, {Muy poco, Poco, Medio, Bastante, Mucho}.
- **Escala intervalo.** En esta escala la distancia entre intervalos es conocida y siempre la misma, si bien no tienen un valor inicial de referencia o cero absoluto.
  - P.e. la temperatura en Celsius donde la diferencia entre  $12^{\circ}\text{C}$  y  $13^{\circ}\text{C}$  es la misma que entre  $24^{\circ}\text{C}$  y  $25^{\circ}\text{C}$ , pero no podemos decir que a  $24^{\circ}\text{C}$  haga ``el doble de calor'' que a  $12^{\circ}\text{C}$ . Fechas, podemos decir que un proyecto estará en el día 200, pero no que un proyecto va a empezar el ``doble de tarde'' que otro.
- **Escala de ratio.** Tienen un valor inicial de referencia o cero absoluto, y permiten definir ratios coherentes con los valores de la escala (se pueden comparar los valores estableciendo proporciones). P.e., temperatura en Kelvin.
- **Escala absoluta.** Las escalas absolutas tienen las características de las escalas anteriores, si bien consisten simplemente en el conteo. P.e., número de programadores.

# Operaciones estadísticas y matemáticas permitidas

Escala	Operaciones estadísticas	Operaciones matemáticas
Nominal	moda	=
Ordinal	mediana	>, <
Intervalo	Media, desviación estándar	+, -
Ratio	Media geométrica, Coeficiente de variación	+, -, ×, ÷

# Evaluación empírica

# Tipos de evaluación empírica

- Las **encuestas** son estudios retrospectivos cuya intención es documentar relaciones y resultados, siendo una de las herramientas más útiles para recabar un buen número de datos que posteriormente serán analizados.
- Los **Casos de estudio** son empleados para identificar y documentar factores clave que pueden afectar los resultados de una actividad.
- **Experimentos formales** se emplean para, de forma controlada y rigurosa, investigar cuantitativamente aquellos factores que afectan a las actividades a realizar.

# Encuestas

- Las encuestas son un método que se emplea para recabar datos que están en la memoria de los entrevistados.
  - Por ejemplo, puede evaluarse un nuevo método o herramienta en una organización a través de un cuestionario de los empleados.
- Las encuestas forman parte de lo que se conoce como investigación a gran escala (*research in-the-large*), pues sirven para recopilar un gran número de datos acerca de diferentes personas y proyectos.
  - Esto además permite que las conclusiones puedan generalizarse, siempre y cuando la selección de los entrevistados haya sido realmente aleatoria y significativa.
- Una vez establecido el objeto(s) de interés, y dependiendo tanto de la población seleccionada como del porcentaje de respuestas esperadas, las entrevistas pueden diseñarse para ser realizadas cara a cara, por correo o a través de Internet (Web o email).

# Tipos de encuestas

- Dependiendo del método utilizado existen tres maneras de llevar a cabo las entrevistas:
  - **Entrevistas estructuradas**, en las que las preguntas corren a cargo del entrevistador. Dichas preguntas están fijadas de antemano, son las mismas para todos los entrevistados y no se modifican durante la entrevista.
  - **Entrevistas no estructuradas**, donde el entrevistado puede ser la fuente tanto de respuestas como de preguntas. El objetivo es obtener la mayor información posible del entrevistado.
  - **Entrevistas semi-estructuradas**, formadas por un conjunto de preguntas abiertas que facilitan el que el entrevistado pueda ofrecer información no prevista por el entrevistador.

# Objetivo de las encuestas

- Pueden utilizarse tanto para generar teorías como para confirmarlas.
  - La **generación de teorías** se realiza extrayendo de las notas de las entrevistas que permitan ``afirmaciones'' avaladas por los datos, es decir, son estudios descriptivos o explicativos.
  - Por otro lado, las entrevistas ayudan a la **formulación de hipótesis**, es decir como estudios exploratorios que ayudan a buscar futuras líneas de investigación.
- Existen guías tanto para la correcta planificación y realización de los cuestionarios como para su análisis. En éste último caso, pueden utilizarse técnicas estadísticas si lo que se desea evaluar son los resultados de cuestionarios cuantitativos.
- Para datos cualitativos y generación de teorías pueden utilizarse por ejemplo el *método de comparación constante y teoría fundamentada*.
- Generalmente las encuestas son cualitativas, aunque pueden ser cuantitativas dependiendo de cómo se haya diseñado la encuesta.

# Teoría fundamentada (*Grounded Theory*)

- Principalmente una técnica cualitativa, que a partir de observaciones o entrevistas, que siguiendo un proceso permiten definir teorías a basándose en anotaciones o códigos (*codes*).
- que consiste en añadir códigos a los resultados, agrupar la información según los códigos y sintetizar las conclusiones..
- Más información:
  - [http://en.wikipedia.org/wiki/Grounded\\_theory](http://en.wikipedia.org/wiki/Grounded_theory)
  - <http://www.groundedtheory.com/>
  - <http://www.scu.edu.au/schools/gcm/ar/arp/grounded.html>

# Casos de estudio

- Los casos de estudio constituyen una técnica de investigación que se basa en la observación de entornos reales para contrastar o validar datos, encontrar relaciones y descubrir tendencias.
- Por ejemplo, en la ingeniería del software se usan mucho para comparar métodos o herramientas, pudiendo ser tanto cuantitativos como cualitativos.
  - Si se desea introducir una nueva herramienta en una organización pero no se está seguro de si el incremento en la productividad merecerá la pena el esfuerzo en inversión, podría diseñarse un caso de estudio como parte de un proyecto piloto cuyo único objeto fuese comparar la productividad de la nueva herramienta con la de la herramienta actual.

# Casos de estudio

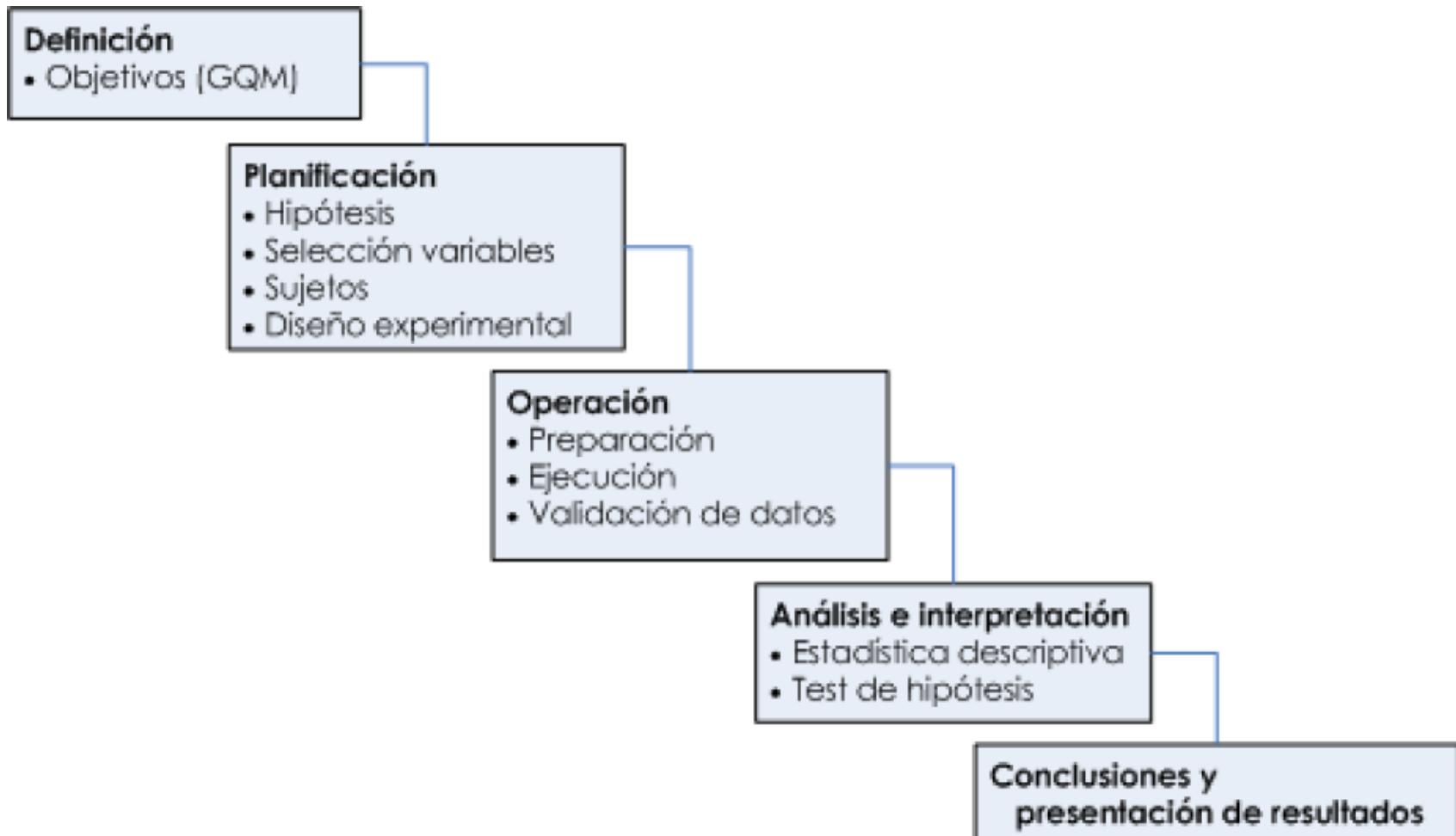
- Los casos de estudio se clasifican como “investigación en lo típico” (*research in the typical*) pues reflejan preferentemente la información de un proyecto representativo en lugar de toda la casuística con la que nos podríamos encontrar.
- La diferencia entre los casos de estudio y los experimentos formales es que en los experimentos las variables son manipuladas, mientras que en los casos de estudio simplemente se analizan dentro de su contexto.
- Los casos de estudio tienen la ventaja de que son más fáciles de diseñar que los experimentos formales, aunque son difíciles de replicar pues se realizan generalmente en entornos industriales concretos.
  - Al no tener los investigadores ningún control sobre las variables, el establecimiento de relaciones causales entre ellas es más difícil y por tanto la generalización de las conclusiones más complicada.

# Experimentación formal

- La experimentación formal busca **medir relaciones causales** con la mayor exactitud posible, i.e., establecer el grado de influencia de unas variables en otras.
  - Para ello es necesario configurar entornos en los que se tenga un alto nivel de control sobre las variables para, modificándolas, medir sus efectos.
  - Fenton y Plfeeger clasifican la experimentación formal como investigación a pequeña escala (*research in the small*) debido a la dificultad de controlar todos los posibles factores.
  - Se suelen realizar en entornos académicos o en organizaciones con un grupo reducido de trabajadores, precisamente por la dificultad mencionada, lo que ha hecho que a estos estudios también se los conozca como experimentos *in vitro*.
- Los experimentos formales son por tanto más difíciles de diseñar y más costosos que las encuestas o los casos de estudio.
- La experimentación formal ayuda a confirmar teorías, explorar relaciones causa-efecto entre variables, corroborar o rechazar creencias sobre métodos, procesos y tecnologías, validar métricas, etc.

# Diseño de experimentos formales

# Proceso y diseño de experimentos



# Proceso en la experimentación formal

- **Definición:** intención del experimento, es decir, cuáles son sus objetivos, en qué contexto se llevará a cabo, cuál es su propósito, etc.
- **Planificación:** Se diseña el experimento mediante el enunciado de las hipótesis, la selección de variables de estudio, la selección de los sujetos y finalmente el diseño del experimento propiamente dicho.
- **Operación:** es la fase de preparación del experimento, que consiste en realizar la selección de participantes, formularios y guías para que los participantes entiendan el proceso, entrenar a los participantes si es necesario, y acciones necesarias para poner en marcha el experimento. Suele realizarse un ***estudio piloto*** para encontrar errores en el diseño.
- **Interpretación:** A la hora de analizar e interpretar los resultados, el tipo de análisis está condicionado por la hipótesis del experimento, el tipo de variables seleccionadas según su escala y el diseño experimental. El primer paso es caracterizar los datos utilizando estadística descriptiva, para posteriormente realizar el ***test de hipótesis*** con el fin de confirmar o rechazar las hipótesis enunciadas.
- **Conclusiones y presentación de resultados:** Informe comentado del proceso seguido. En este informe se debe enfatizar por qué es relevante la hipótesis del estudio y cómo se han solventado o minimizado las amenazas a la validez e información que permita crear posibles réplicas del experimento.

# Hipótesis

- Se plantean dos tipos de hipótesis:
  - **Hipótesis Nula ( $H_o$ )**: no hay diferencia entre dos o más tratamientos (es decir, entre dos o más métodos, herramientas u otras condiciones cuyos efectos se quieran medir) con respecto a la variable dependiente que se mide (por ejemplo, la calidad, la eficacia o la productividad).
  - **Hipótesis alternativa o experimental ( $H_a$ )**: sí hay diferencia significativa.

# Validación de la hipótesis

- Se supone que la hipótesis nula es cierta.
- Para decidir si hay o no una diferencia estadísticamente significativa entre dos o más tratamientos, es necesario contrastar la hipótesis nula y fijar un nivel de confianza
  - sus valores más frecuentes suelen ser 0,95 y 0,99.
- Hasta que no se acaba de analizar todos los datos y de aplicar el test estadístico adecuado, no se sabe si se acepta o se rechaza la hipótesis nula con un nivel de significación (0,05 o 0,01).
- Las reglas de decisión del contraste establecen que el rechazo de  $H_0$  implica la aceptación de  $H_a$  y, recíprocamente, si se acepta  $H_0$  se rechaza  $H_a$ .

# Tipos de errores y Potencia

- **Error de tipo I** consiste en rechazar incorrectamente  $H_0$ , está bajo control a través del nivel de significación, también llamado  $\alpha$ .
- **Error de tipo II** consiste aceptar  $H_0$  siendo falsa, se controla principalmente mediante el diseño experimental.
- **Potencia del contraste estadístico**, definida por  $1-\beta$ , que mide la probabilidad de rechazo de la hipótesis nula cuando realmente es falsa. Por eso no hay que omitir los dos tipos de errores que se pueden cometer en una investigación empírica.
  - De aquí se deduce que si la potencia depende de  $\alpha$ , del tamaño de la muestra y del tamaño del efecto y, además, estos dos últimos factores se consideran constantes en el diseño, es posible aumentar la potencia del contraste estadístico o disminuir el riesgo  $\beta$  a expensas de incrementar  $\alpha$ .

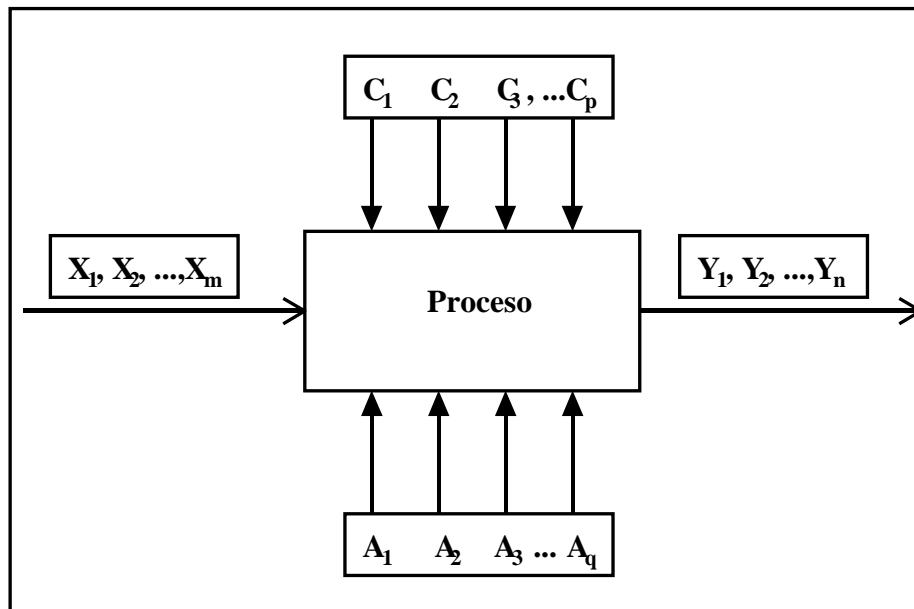
# Tratamientos, unidades y sujetos experimentales

- El siguiente paso es concretar la hipótesis planteada, lo que lleva a identificar tanto los tratamientos como las unidades y los sujetos experimentales.
- Un experimento tiene por objeto descubrir las relaciones existentes entre las variables, fundamentalmente, determinar el efecto de la variable independiente, también denominada factor, sobre la dependiente.
- Cada factor se divide en un número determinado de niveles. En el caso de que el estudio se lleve a cabo con dos variables independientes, la combinación del nivel de un factor con el nivel de otro constituye un tratamiento.
  - Por esta razón, cuando se trata de una sola variable independiente, el término “tratamientos” equivale a “niveles”.
  - Así pues, las unidades y los sujetos experimentales son los objetos materiales y las personas participantes, a los que se aplica un tratamiento.

# Variables

- Las variables determinan los tipos de análisis estadísticos que se pueden aplicar (por ejemplo, el tipo de escala)
- Las variables en un experimento formal son la variable independiente y la dependiente, estableciendo la hipótesis de que la variable independiente influye en la dependiente.
- Existen otras de gran importancia que también representan una “fuente de variación” en tal variable [Kish, 1995]:
  - Controlada (*controlled*): es una variable extraña que puede ser adecuadamente controlada mediante el diseño experimental.
  - Enmascarada (*confounded variable*): es una variable extraña no controlada que tiene la propiedad de variar simultáneamente con los cambios en los niveles de la variable independiente. Esta circunstancia hace que rivalice con la variable independiente como posible causa.
  - Aleatoria (*randomized*): es una variable extraña no controlada que se trata como un error aleatorio. La aleatorización se puede considerar como una forma de control experimental, si bien es distinta de las que se aplican en las variables controladas.

- Un experimento constituye un proceso con variables independientes ( $X_1, X_2, \dots, X_m$ ) tan poderosas como sea posible, gracias a la aplicación de un control interno casi perfecto, y con variables dependientes ( $Y_1, Y_2, \dots, Y_n$ ) lo suficientemente sensibles como para detectar las diferencias.



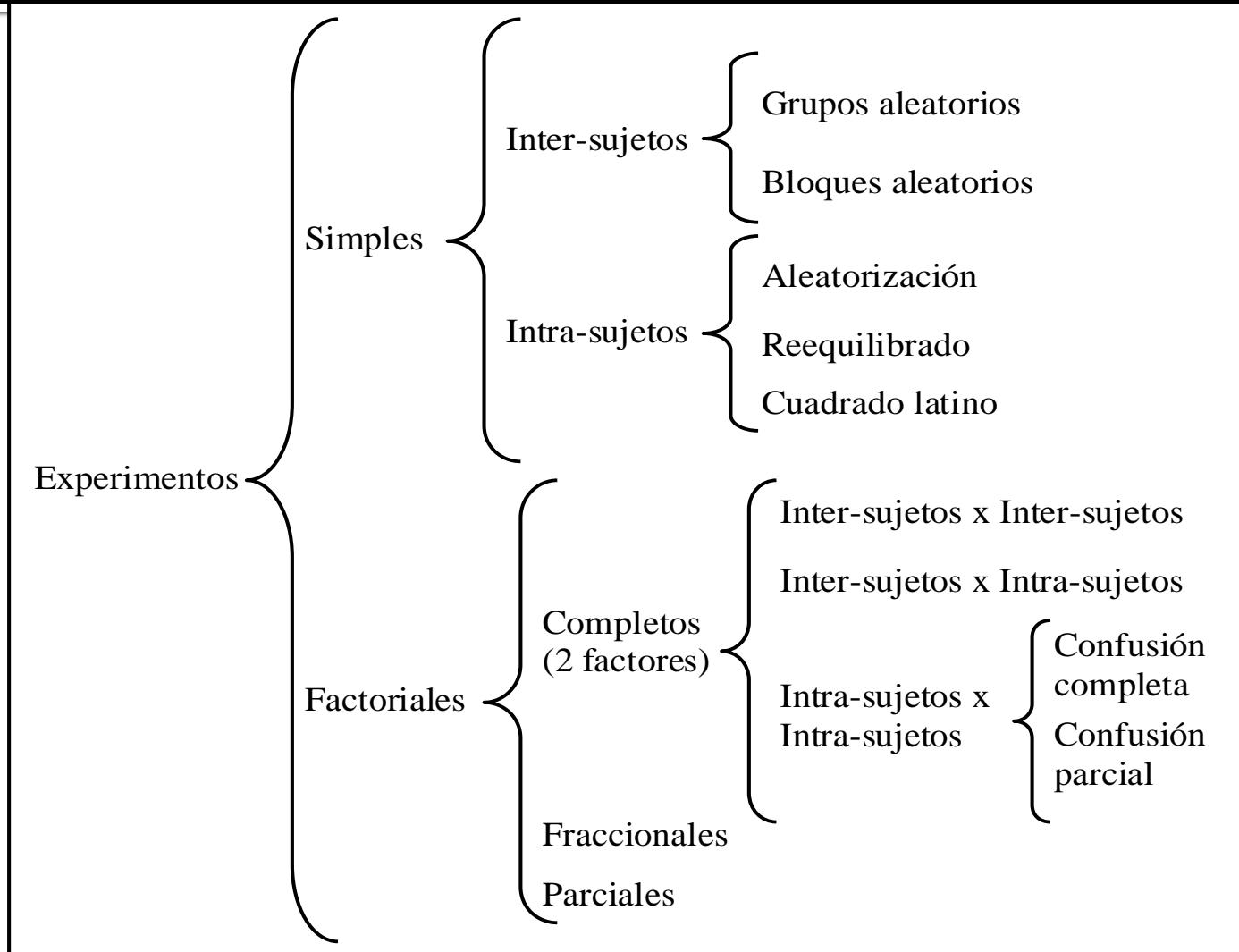
# Propiedades

- **Fiabilidad.** Se dice que un experimento es fiable cuando, al repetirse, se obtienen los mismos resultados. Para ello es preciso el control y la replicación, interna o externa, del experimento.
- **Validez constructiva.** Es el grado con que, tanto las variables independientes como las dependientes, reflejan o miden con certeza las hipótesis del experimento.
- **Validez interna.** Es el grado de seguridad con el que se puede atribuir que los cambios de la variable dependiente se deben a la influencia de la variable independiente que se estudia.
- **Validez externa.** Es el poder de generalización de los resultados obtenidos en el laboratorio a las condiciones normales.

# Amenazas de la validez interna

- Los efectos de la **selección** tienen su origen en las variaciones naturales en el rendimiento de cada sujeto. Algunos ejemplos son la experiencia previa al experimento programando en un cierto lenguaje, la inteligencia o la habilidad.
- Los efectos de la **maduración** son causados por el paso del tiempo, concretamente, por el cansancio o por el aprendizaje de los sujetos según avanza el experimento.
- Los efectos de la **historia** se deben a eventos específicos que podrían ocurrir en el transcurso de un experimento. Es un ejemplo el caso del investigador que ha enfermado y al que hay que reemplazar, un fallo en el ordenador o cualquier otra interrupción no deseada.
- Los efectos de la **instrumentación** pueden tener su origen en las diferencias en los materiales experimentales que se emplean.
- Los efectos de la **mortalidad** son debidos a la pérdida de sujetos a lo largo de un experimento. Puede servir como ejemplo la falta de motivación, originada por el trabajo reiterativo, y su larga duración.
- La causa de los efectos de la **práctica** o del **orden de presentación** es la cantidad de práctica realizada antes de la presentación correspondiente a un tratamiento determinado.

# Taxonomía de los experimentos



# Diseños simples – Inter-sujetos

## Grupos aleatorios

- Grupos aleatorios o también denominada grupos independientes
- A cada sujeto se le asigna un único tratamiento
- Los grupos de sujetos correspondientes a los distintos niveles de la variable independiente se forman mediante asignación aleatoria.
- Con esta técnica se controla, por equilibrado, las variables extrañas que están presentes en los sujetos (experiencia, etc.)
- En general se recomiendan los diseños simples
  - Más fáciles de diseñar y evaluar.

# Diseños simples – Inter-sujetos

## Grupos aleatorios con bloques

- Grupos aleatorios con bloques o grupos relacionados
- La técnica de control con la que se consiguen grupos emparejados (*matched*) se denomina bloqueo, ya que se forman bloques en una variable relacionada con la variable dependiente.
- Por ejemplo, Tryggeseth (1997) aplica esta técnica e un experimento para evaluar la función de la documentación en el mantenimiento del software.
  - Antes de ejecutarlo, los 34 estudiantes se dividen en dos grupos, pero no aleatoriamente. Con el fin de asegurar una distribución homogénea de ambos grupos respecto a la variable habilidad con el lenguaje C++, se procede así:
    - Se mide a los participantes en la variable extraña que se quiere controlar, es decir, la habilidad con el lenguaje C++, mediante un test.
    - Se agrupa a los sujetos en cuatro bloques con valores similares en dicha variable.
    - Se asigna al azar a los sujetos de cada bloque a uno de los dos grupos del experimento, hasta que ambos tengan el mismo tamaño.

Tryggeseth, E., "Report from an experiment: impact of documentation on maintenance", Empirical Software Engineering, vol. 2 (2) 1997, pp. 201-207

# Diseños simples – intra-sujetos

- A cada sujeto se le asignan todos los tratamientos. En lugar de intentar que los distintos grupos sean lo más parecidos posibles, una fórmula mejor consiste en que los grupos asignados a los distintos tratamientos estén formados por los mismos sujetos.
- Ventajas:
  - se garantiza el control de todas las variables achacables a las diferencias entre los sujetos que pudieran contaminar los resultados del experimento;
  - Se ahorra esfuerzo, pues se consigue la misma información con un número menor de sujetos. A este tipo de diseños se les denomina diseños con los mismos sujetos o diseños de medidas repetidas.
  - Los diseños de medidas repetidas reducen las amenazas a la validez interna debidas a las diferencias individuales.
- Pero , se pueden dar efectos distorsionantes por la maduración, la instrumentación, la mortalidad y la práctica se tienen que controlar.
  - Un modo de amortiguar el efecto del aprendizaje consiste en que todos los sujetos participen en todos los tratamientos de forma rotatoria y en días diferentes para contrarrestar el cansancio.
- Sin embargo, pese a los esfuerzos por controlarlos, siempre está presente el efecto de la práctica (orden de los tratamientos) , que se controla y/o se mide con un adecuado plan en cuanto al orden de presentación de los niveles de la variable independiente. Las estrategias con las que se controla son:
  - Aleatorización
  - Reequilibrado (*counter-balancing*) o cuadrado latino (*latin square*) dependiendo si los niveles de la variable independiente son dos , o mayor o igual a 3 respectivamente, que además permiten medir el efecto de la práctica.

# Diseño simple – Intra-sujetos

## Control mediante aleatorización

- Cuando no interesa conocer la cuantía del efecto de la práctica, es suficiente para su control una estrategia de aleatorización: simple o por bloques.
  - Estas dos técnicas ofrecen una cierta garantía de que un posible efecto del orden de presentación queda amortiguado.
  - Este hecho resulta más evidente en la aleatorización por bloques, puesto que la amenaza del orden de presentación se reparte de forma bastante homogénea entre los diferentes tratamientos experimentales

# Diseño simple – Intra-sujetos Reequilibrado

- La técnica del reequilibrado consiste en repetir los dos tratamientos experimentales, *A* y *B*, de tal forma que primero se presentan en un orden y, después, en el inverso.
- Si el primer tratamiento afecta al segundo, produciendo un desequilibrio, se anula, al invertirse este efecto en la segunda repetición, y queda reequilibrado.
  - Se habla de **reequilibrado incompleto** cuando el control del orden de presentación abarca a todo el colectivo; es decir, los tratamientos se aplican en el orden (*AB*) a un grupo y en el orden inverso (*BA*) al otro.
  - En cambio, se dice que es **reequilibrado completo** cuando la secuencia *ABBA* se aplica a todos los sujetos.

# Diseño simple – Intra-sujetos Reequilibrado. Ejemplo

- Daly *et al.* (1996) describe 2 experimentos ideados para evaluar la profundidad de herencia, cuya única diferencia es el nivel de jerarquía: 3 en uno y 5 en el otro.
  - El diseño usado es de medidas repetidas con bloques aleatorios
  - Todos los sujetos se miden con respecto a los dos niveles de la variable independiente, y se emparejan sobre la base de su conocimiento en programación orientada a objetos, para su posterior asignación al azar a uno de los dos grupos: A o B.
  - En estos estudios empíricos, la variable experimental es la profundidad de la herencia: nivel máximo (3 o 5) o nivel 1 (no tiene jerarquía). Para amortiguar el efecto del aprendizaje se construyen dos programas, mientras que para controlar el efecto de la práctica se emplea la técnica de reequilibrado incompleto.

	Sesión 1	Sesión 2
Grupo A (16)	Prog1conH	Prog2sinH
Grupo B (15)	Prog1sinH	Prog2conH

# Diseño simple – Intra-sujetos

## Cuadrado latino

- La lógica implícita en la técnica de control mediante con dos niveles en la variable independiente es compensar los efectos de la práctica aplicando dos tratamientos en los dos órdenes posibles:  $AB$  y  $BA$ .
- Pero si la variable independiente tuviese 3 niveles, habría que aplicar esos niveles en 6 secuencias diferentes:  $ABC$ ,  $ACB$ ,  $BAC$ ,  $BCA$ ,  $CAB$  y  $CBA$ .
  - Como esto requiere un gran esfuerzo, se han desarrollado técnicas para controlar el posible efecto de la práctica cuando se presentan más de dos tratamientos.
- Estas técnicas permiten los denominados diseños de cuadrado latino.
  - En este tipo de diseños el control del efecto de la práctica se lleva a cabo seleccionando un grupo de secuencias de presentación entre todas las posibles. Se eligen tantas secuencias como tratamientos haya en el experimento, de tal modo que cada tratamiento sólo ocurra una vez en cada posición dentro del conjunto de secuencias elegidas. Por ejemplo los tres órdenes de presentación para el caso de tres niveles seleccionados entre las seis permutaciones mencionadas.

B	A	C
C	B	A
A	C	B

# Diseños complejos o factoriales

- Los diseños factoriales se caracterizan por estudiar dos o más variables independientes ( $vi$ ), denominadas factores que, a su vez, se dividen en niveles.
- Una primera clasificación de diseños factoriales (por ejemplo,  $2 \times 2$  o  $2^2$ ) distingue entre diseños cruzados y anidados:
  - **Diseños cruzados.** Todos los niveles de una  $vi$  (factor A) se combinan con los niveles de la otra (factor B). Por eso, es posible examinar la influencia de la interacción  $A \times B$  sobre la variable dependiente. Si no hay interacción, entonces hay que explorar los efectos principales debidos a los factores A y B.
  - **Diseños anidados.** Los niveles del factor B están anidados dentro de los niveles del factor A. Es imposible estudiar la interacción  $A \times B$ , y el efecto principal del factor B sólo se puede explorar dentro de cada nivel del factor A.

# Diseños cruzado vs. anidados

		Factor B	
		b <sub>1</sub>	b <sub>2</sub>
Factor A	a <sub>1</sub>	a <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>
	a <sub>2</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>2</sub>

Diseño cruzado

Factor A		Factor B	
a <sub>1</sub>		a <sub>2</sub>	
Factor B		Factor B	
b <sub>1</sub>	b <sub>2</sub>	b <sub>1</sub>	b <sub>2</sub>
a <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>2</sub>

Diseño anidado

# Replicación de experimentos

- Asimismo, es importante lograr la fiabilidad de un experimento mediante su replicación, ya sea interna o externa. Para llevar a la práctica una réplica externa es preciso crear paquetes de laboratorio en los que se informe sobre el diseño, los instrumentos y los procesos utilizados en el experimento, así como sobre los métodos empleados durante el análisis. Según [Basili *et al.*, 1999], los tipos de replicaciones se agrupan en tres categorías:

# Tipos de réplicas

- Réplicas que no hacen variar ninguna de las hipótesis de la investigación, o sea, que no modifican ninguna de las variables independientes y dependientes del experimento original.
  - Réplicas estrictas.
  - Réplicas que varían la manera en que se ejecuta el experimento.
- Réplicas que varían las hipótesis de la investigación.
  - Réplicas que cambian las variables independientes.
  - Réplicas que cambian las variables dependientes.
  - Réplicas que cambian las variables de contexto en el entorno en que evaluó la solución. Así, ayuda a mejorar la validez externa la recurrencia a profesionales en lugar de estudiantes en una réplica con los mismos procesos y productos para determinar si se llega a los mismos resultados.
- Réplicas que amplían la teoría; es decir, se realizan varios cambios para comprobar si los resultados permanecen.

# Guía para la investigación empírica

- Kitchenham et al han definido una guía para tener en cuenta aspectos de investigación empírica en la ingeniería del software, aunque generalizable (de hecho, se basan en guías médicas).
  - Provee guías en contexto, diseño, recabar datos, análisis y presentación de los resultados.

Kitchenham, B.A.; Pfleeger, S.L.; Pickard, L.M.; Jones, P.W.; Hoaglin, D.C.; El Emam, K.; Rosenberg, J., "Preliminary guidelines for empirical research in software engineering," *Software Engineering, IEEE Transactions on*, vol.28, no.8, pp. 721-734, Aug 2002.  
Digital Object Identifier 10.1109/TSE.2002.1027796

<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1027796&isnumber=22077>