

Módulo II. Diseño y evaluación en la experimentación formal.

Introducción a las técnicas de investigación.

Contenidos

- Diseño y evaluación en la experimentación formal.
 - Escalas y medición
 - Proceso y diseño de la experimentación formal

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

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
 - Dolado Cosín, J.J. y Fernández Sanz L. (2000). Medición para la gestión en la Ingeniería del Software. Ra-Ma.
 - 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. Moreno, A., (2001) Basics of Software Experimentation, Kluwer Academic Publishers.

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++, Python}.
- **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°C y 13°C es la misma que entre 24°C y 25°C, pero no podemos decir que a 24°C haga "el doble de calor" que a 12°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	+, -, ×, ÷

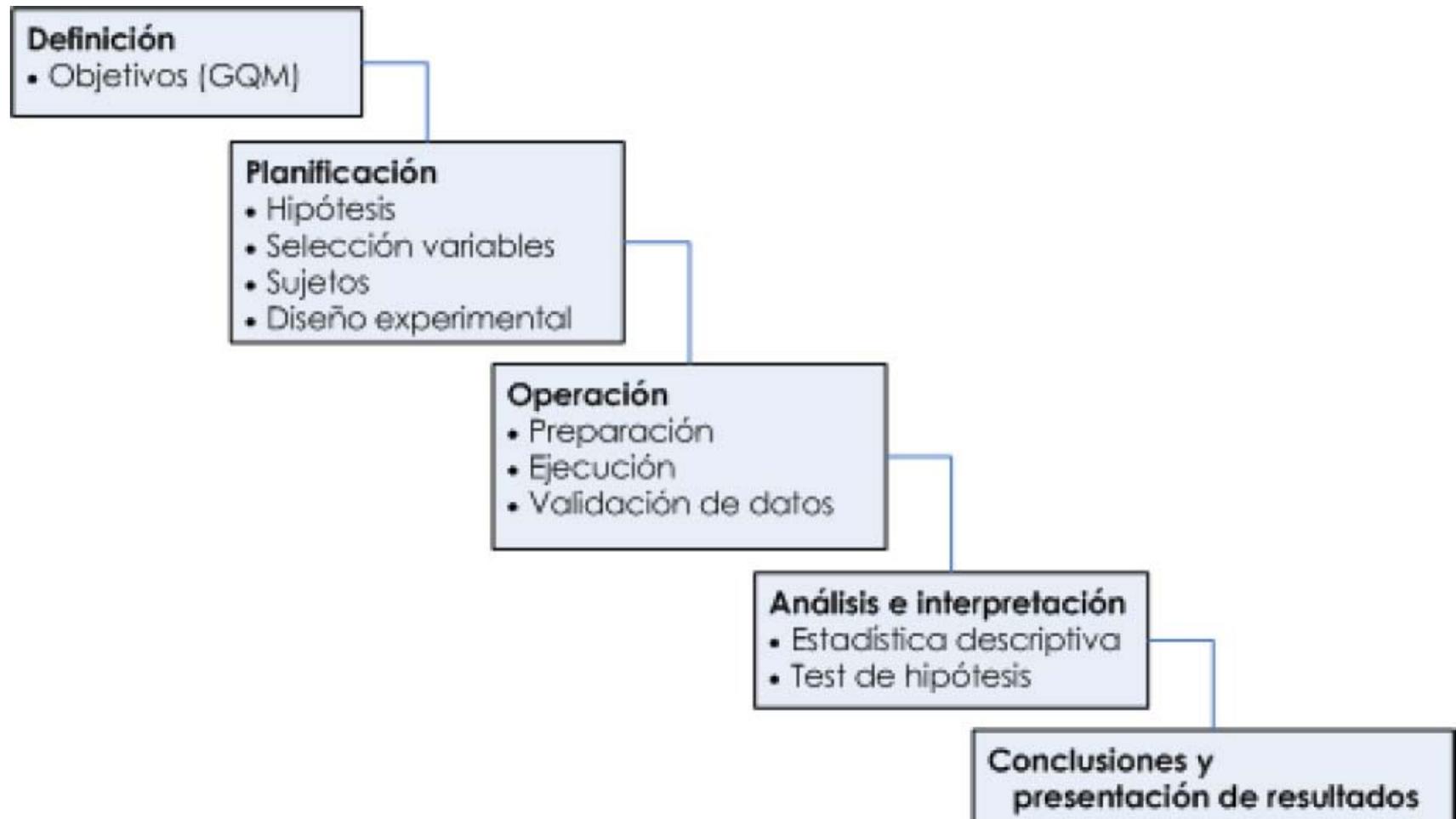
Evaluación empírica

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.
 - Objetivo-Preguntas-Métricas *Goal Question Metric*)
- **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, consistente en 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.

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.

DOI 10.1109/TSE.2002.1027796

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