

Máster en Ingeniería y Tecnología de Sistemas Software

Curso académico 2024/2025

Ingeniería del Software Experimental

Tema 2: Experimentos Controlados

Depto de Sistemas Informáticos y Computación

Universitat Politècnica de València, España

email: sabrahao@dsic.upv.es



UNIVERSIDAD
POLITECNICA
DE VALENCIA

DSIC

Objetivos

- Presentar los **conceptos** relacionados con experimentos en Ingeniería del Software
- Presentar un **proceso experimental** para guiar el diseño y ejecución de experimentos
- Discutir la relevancia de las **replicaciones** de experimentos
- Proporcionar **recomendaciones** sobre cómo realizar experimentos en el ámbito de la Ingeniería del Software



Contenido

- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- Agregación de resultados
- Recomendaciones
- Conclusiones
- Literatura relevante

Experimentos

Conceptos
generales



- Un experimento es una investigación empírica que investiga las **relaciones de causa-efecto**.
- Los experimentos se llevan a cabo cuando el investigador quiere el **control de la situación**, con la manipulación directa, precisa y sistemática del comportamiento del fenómeno a estudiar.
- **Objetivo principal:** probar hipótesis o relaciones entre variables
- **Ventaja:** permiten investigar las situaciones en las que las hipótesis son verdaderas y proporcionan el contexto en el que se recomiendan normas, métodos y herramientas para su uso.



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Tipos de experimentos

Conceptos
generales



■ Orientados a personas:

Un experimento en el cual los individuos o los equipos llevan a cabo **una o más tareas de ingeniería de software** para comparar diferentes poblaciones, procesos, métodos, técnicas, lenguajes o herramientas.

- Por ejemplo, se quiere comparar el desarrollo “tradicional vs. ágil” del software

■ Orientados a la tecnología:

Generalmente se aplican diferentes herramientas a diferentes objetos.

- Por ejemplo, al mismo programa se le aplican dos técnicas de generación de casos de prueba diferentes.



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Experimentos

Conceptos
generales



**Variable independiente
(factor-tratamiento)**



Efecto



**Variable dependiente
(salida)**



**Otras variables independientes
(controladas)**
(experiencia de los sujetos,
complejidad de las tareas y de
los objetos experimentales, etc.)

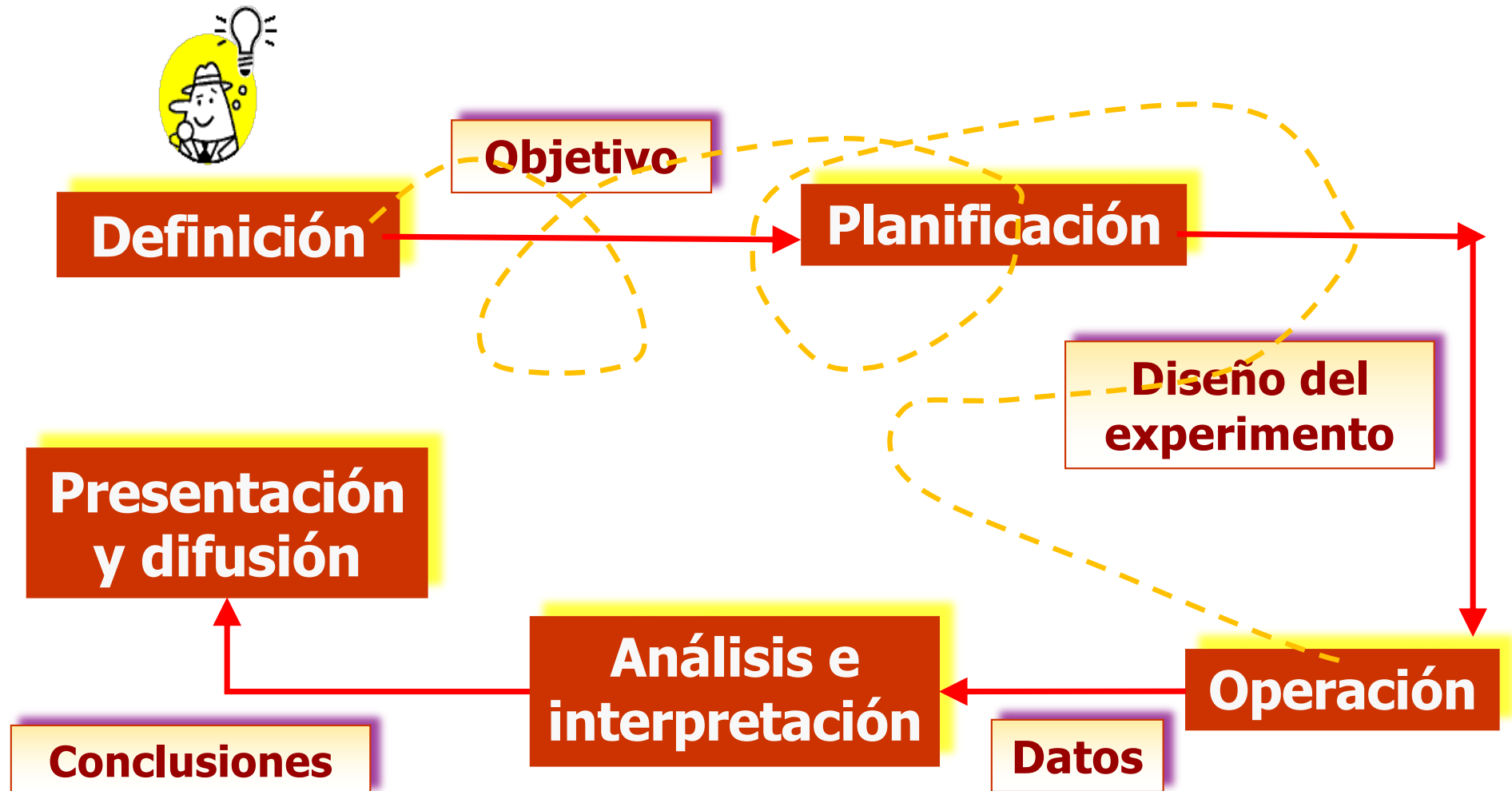


UNIVERSIDAD
POLITECNICA
DE VALENCIA

Contenido

- Experimentos
 - Conceptos generales
- **Proceso experimental**
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- Agregación de resultados
- Recomendaciones
- Conclusiones
- Literatura relevante

Proceso Experimental



(Wohlin et al., 2000-2012)



UNIVERSIDAD
POLITECNICA
DE VALENCIA



Proceso Experimental: Definición

¿Por qué se realiza el experimento?

- Plantilla GQM para definición de objetivos:

Analizar

Método orientado a objetos y estructurado

(¿qué es lo que se estudia?)

con el propósito de

Evaluar

(¿cuál es la intención?)

con respecto a

Productividad

(¿cuál es el efecto estudiado?)

desde el punto de vista del

Investigador

(¿a la vista de quién?)

en el contexto de

Estudiantes de grado

(¿dónde se lleva a cabo el estudio?)



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Proceso Experimental: Definición

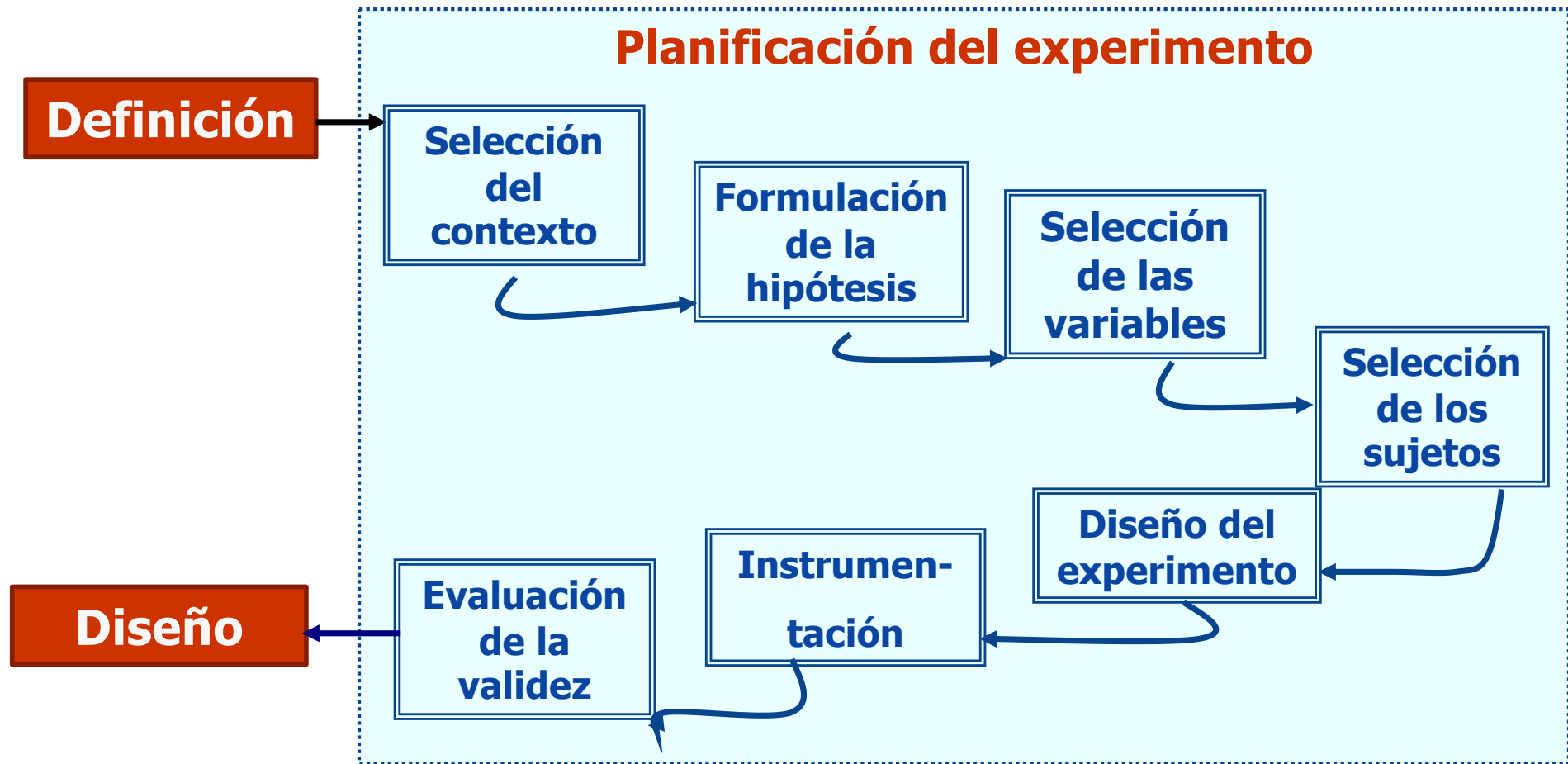


■ Ejemplos de definición de objetivos

Objeto de estudio	Propósito	Enfoque de calidad	Perspectiva	Contexto
Producto	Caracterizar	Efectividad	Desarrollador	Sujetos
Proceso	Supervisar	Coste	Mantenedor	Objetos
Modelo	Evaluar	Fiabilidad	Gestor de proyectos	
Métrica	Predecir	Mantenibilidad	Cliente	
Teoría	Controlar	Portabilidad	Usuario	
	Cambiar		Investigador	



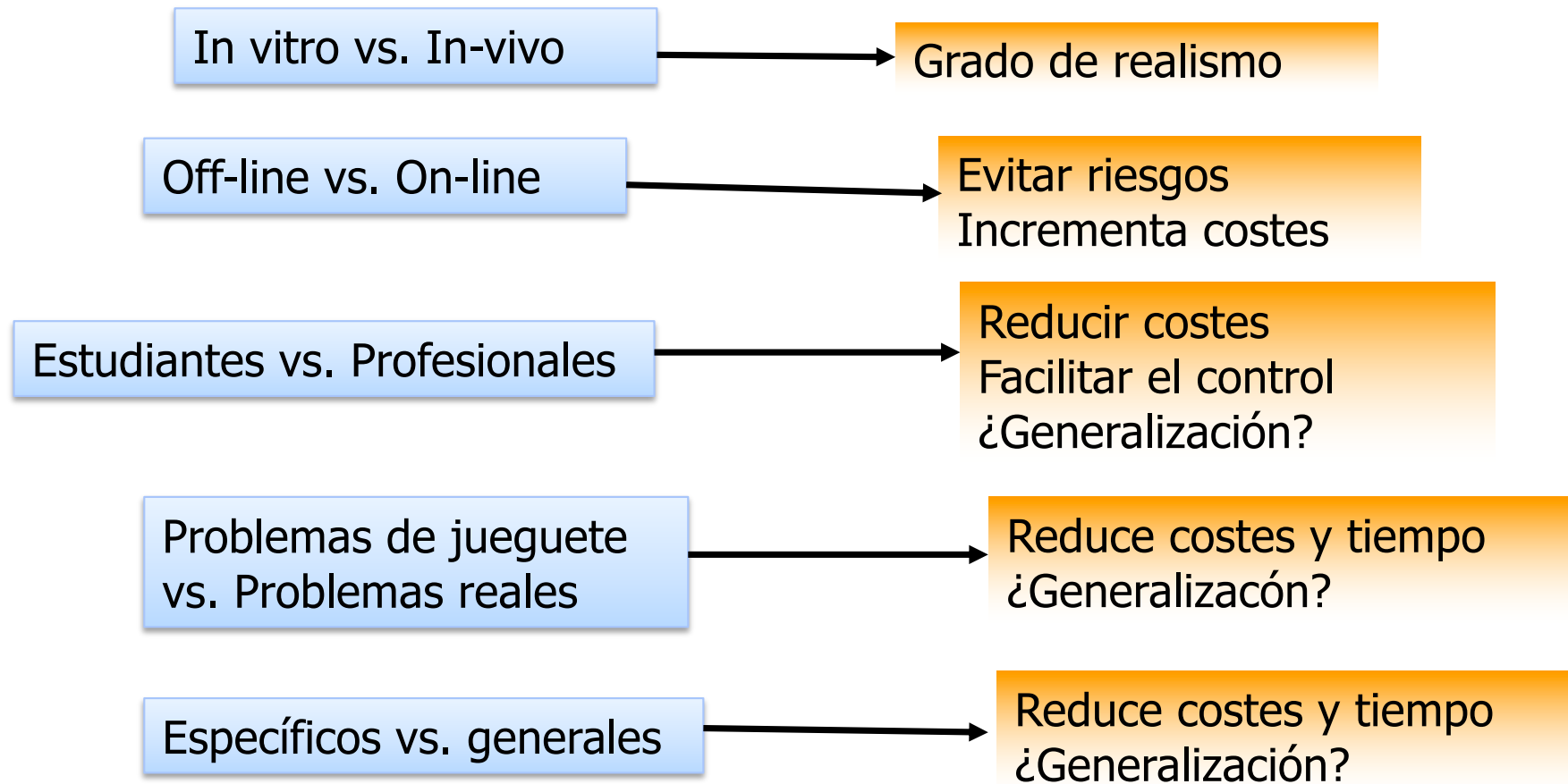
Proceso Experimental: Planificación



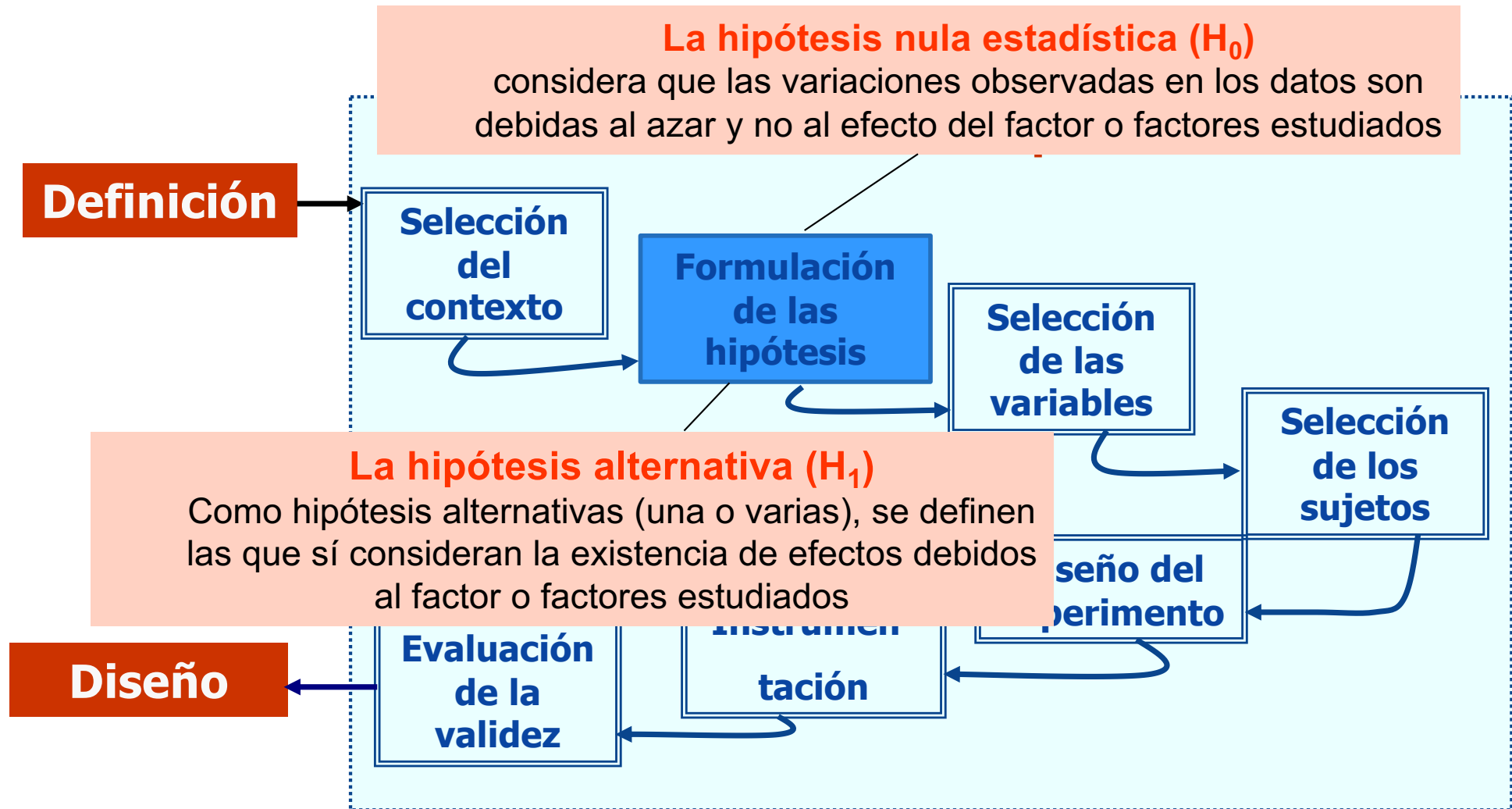
Proceso Experimental: Planificación



Selección del contexto



Proceso Experimental: Planificación



Proceso Experimental: Planificación



Formulación de la hipótesis

- Se deriva del objetivo: una o más hipótesis:

Plantilla GQM

- Analizar *métodos OO y estructurados*
- Con el propósito de *Evaluar* con respecto a su *productividad*
- Desde el punto de vista del *investigador*
- En el contexto de *estudiantes de grado*

H_0 : Productividad OO = Productividad EST

(*hipótesis nula - conservadora, no existe efecto del tratamiento*)

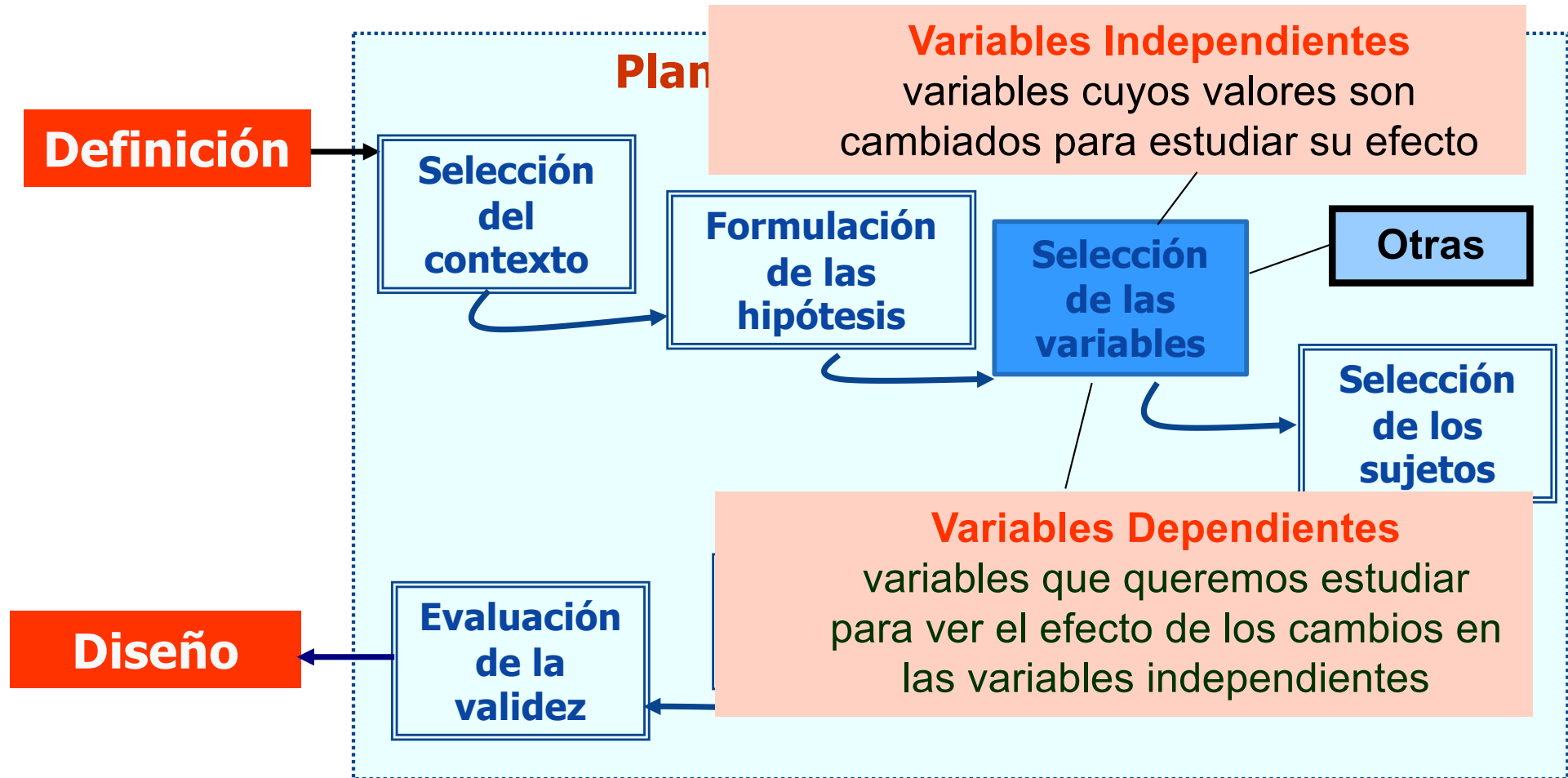
H_1 : Productividad OO \neq Productividad EST

(*hipótesis alternativa - existencia de efectos debido al factor estudiado*)

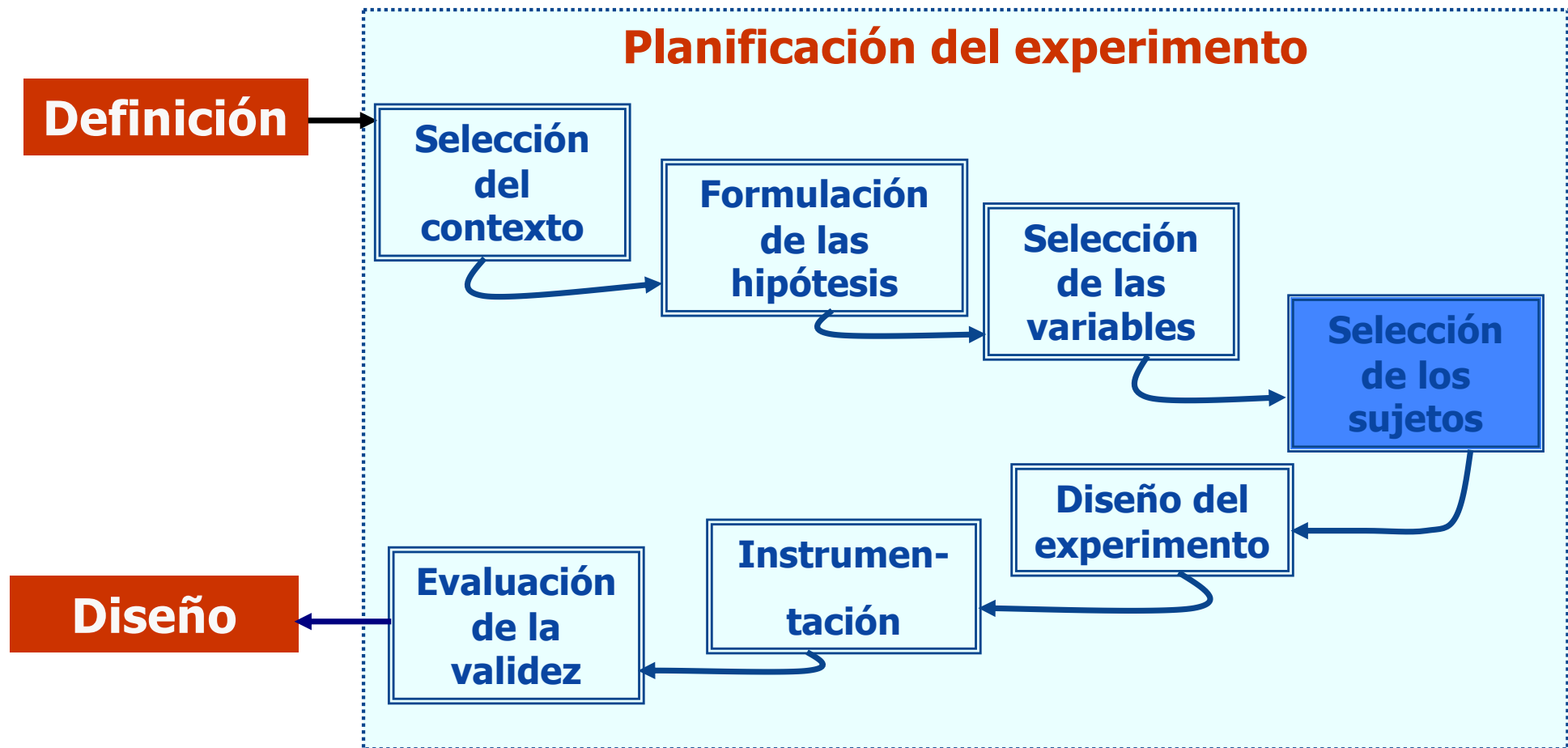


UNIVERSIDAD
POLITECNICA
DE VALENCIA

Proceso Experimental: Planificación



Proceso Experimental: Planificación



Proceso Experimental: Planificación



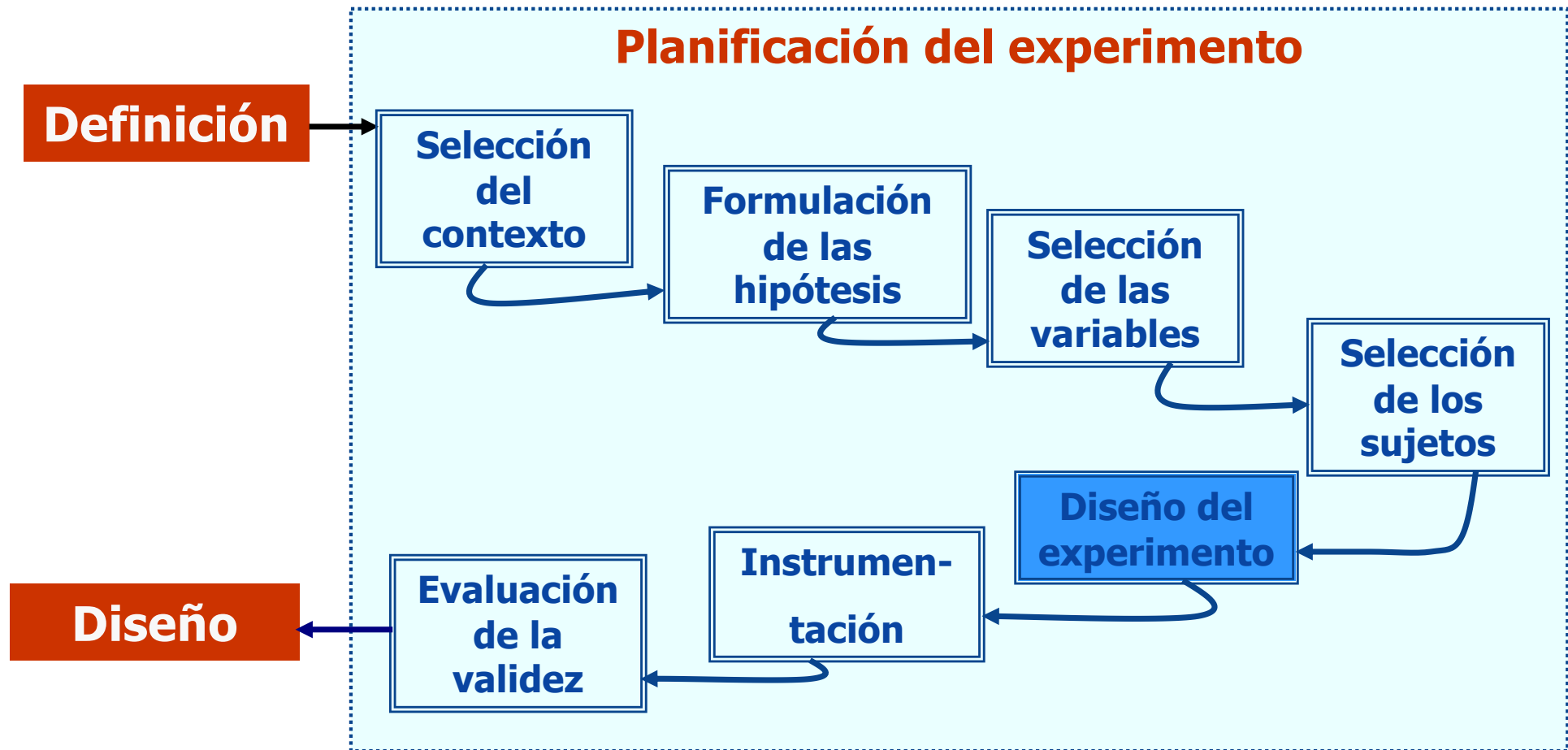
Selección de Participantes

- **¿Cómo seleccionar los Participantes?**
 - **Muestras probabilísticas o no probabilísticas**
 - Muestreo simple aleatorio, sistemático, estratificado aleatorio
 - Muestreo por conveniencia, por cuota
- **¿Y el tamaño de la muestra?**
 - Si hay gran variabilidad en la población se necesita un muestra de mayor tamaño
 - La técnica de análisis puede influenciar la elección del tamaño de la muestra

La muestra debe ser representativa de la población



Proceso Experimental: Planificación



Proceso Experimental: Planificación



Diseño del Experimento

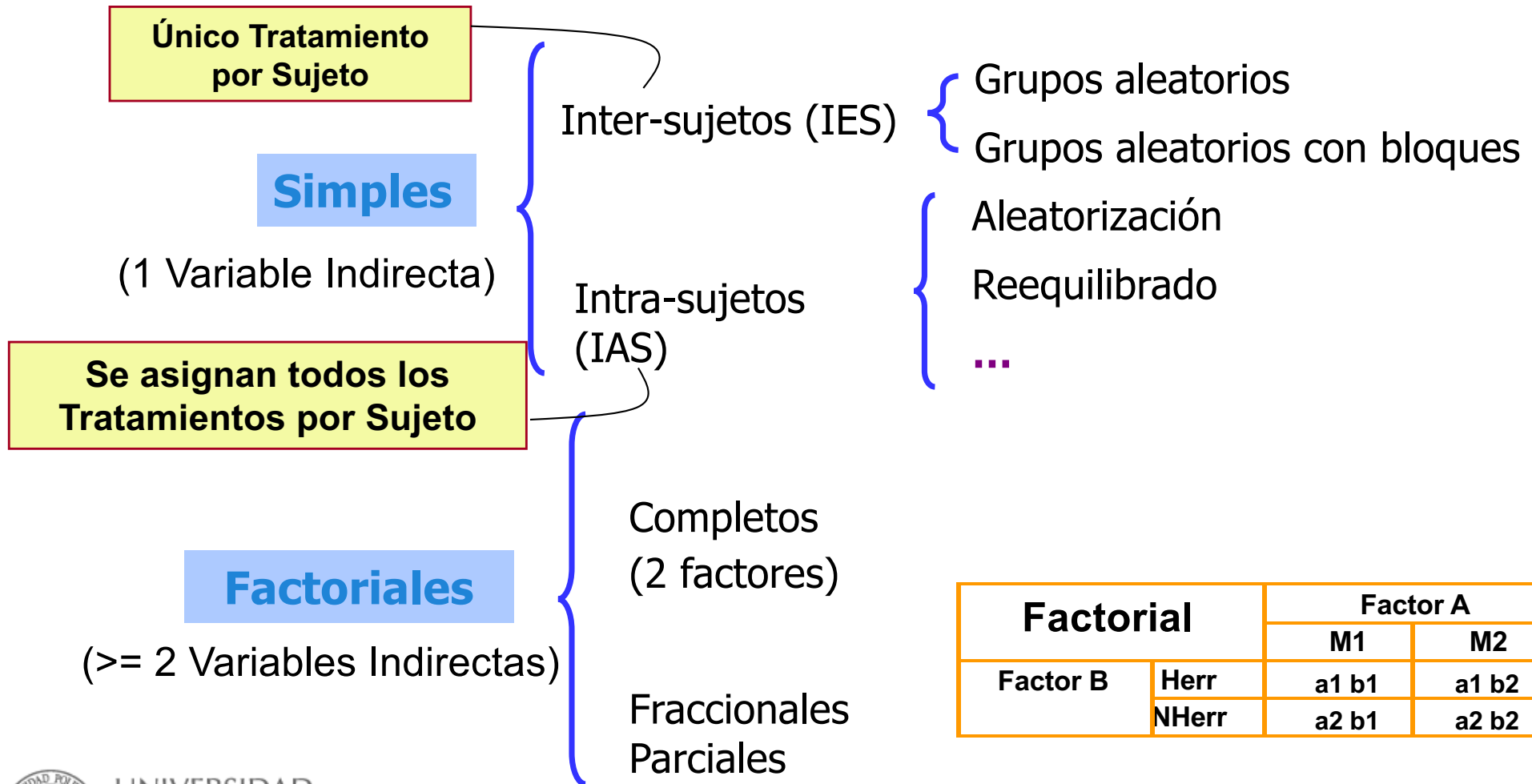
- Describe cómo se organizan los tests
- Relacionado con el análisis, interpretación y conclusiones del experimento
- Preguntas Relevantes:
 - **¿Cuántas variables independientes?**
 - Sólo una → Experimentos simples
 - Más de una → Experimentos factoriales
 - **¿Cuántos tratamientos por sujeto?** **Medidas Repetidas**
 - **¿Cómo controlar los factores extraños?** **Bloqueo / Aleatorización**
- Estas respuestas dependen de las amenazas a la validez que queramos controlar



Proceso Experimental: Planificación

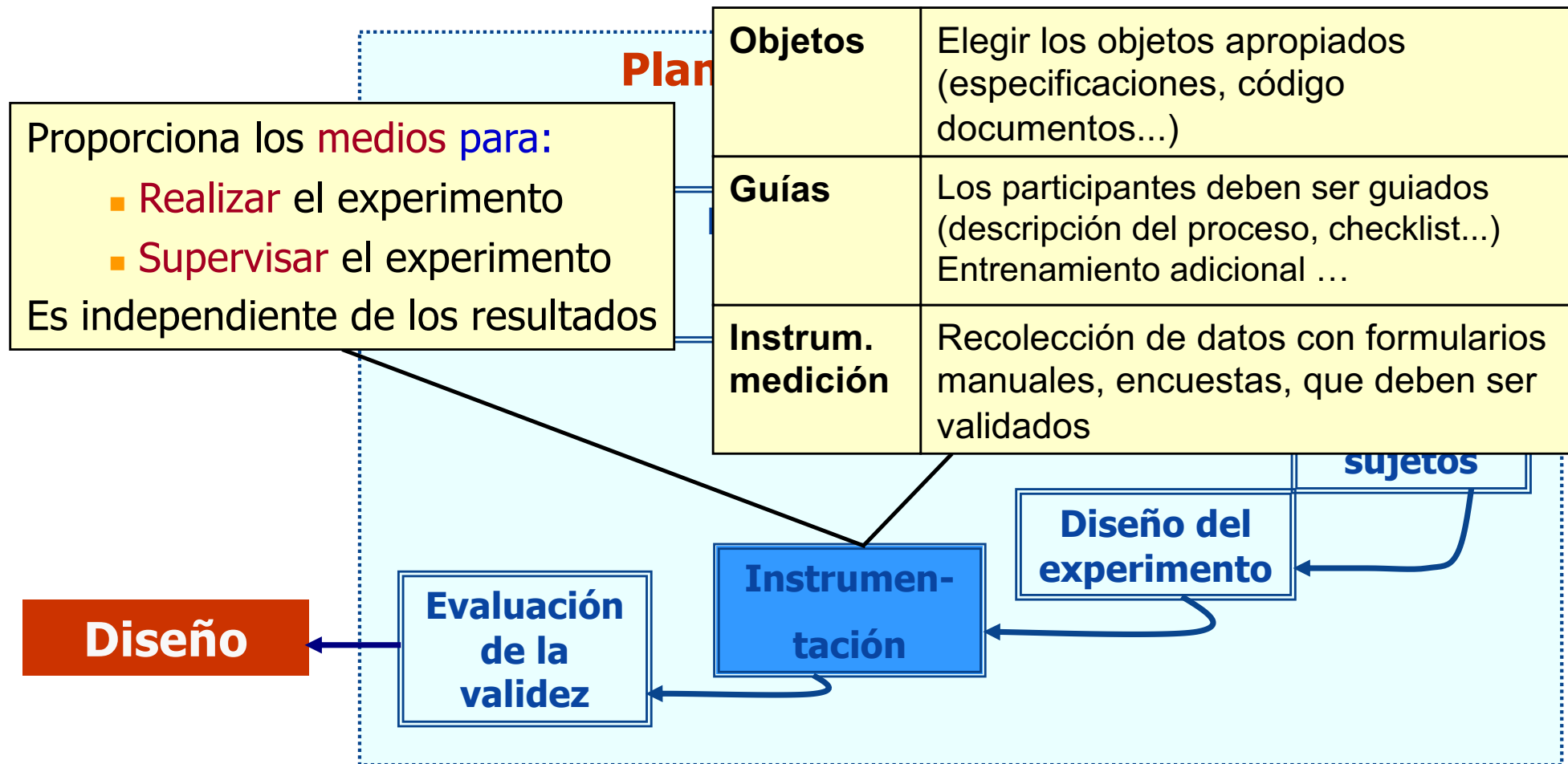


Diseño del Experimento: Taxonomía



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Proceso Experimental: Planificación



Proceso Experimental: Planificación



Proceso Experimental: Planificación



Evaluación de la Validez

- Validez interna: ¿El **tratamiento** causa “realmente” el **efecto**?
- Validez de externa: ¿Pueden **generalizarse** los **resultados** obtenidos?
- Validez de constructo: ¿Hasta qué punto las **medidas seleccionadas** miden las **variables** que aparecen en la **hipótesis**?
- Validez de las conclusiones: ¿Hasta qué punto las **conclusiones** son estadísticamente **válidas**?



Proceso Experimental: Planificación



Amenazas a la validez interna

Historia	Tratamientos diferentes aplicados sobre el mismo objeto en momentos diferentes... ¿las circunstancias son las mismas?
Maduración	Los sujetos reaccionan cuando el tiempo pasa (cansancio, aburrimiento, aprendizaje)
Selección	¿La muestra de la población es representativa para toda la población?
Instrumentación	¿Se han diseñado correctamente los artefactos utilizados en el experimento? Documentos que van a ser inspeccionados ..
Mortalidad	Personas que abandonan el experimento
Orden de presentación	¿Cuánto conoce el sujeto sobre el tratamiento? Se altera para disminuir efectos de aprendizaje.



Proceso Experimental: Planificación



Amenazas a la validez externa

- **Interacción entre selección y tratamiento**
 - La muestra no es representativa de la población
- **Interacción entre el material y el tratamiento**
 - Material no representativo
 - Problemas de juguete, métodos obsoletos
- **Interacción entre la historia y el tratamiento**
 - El experimento se lleva a cabo en un momento especial que afecta a los resultados

Amenazas a la validez de conclusiones

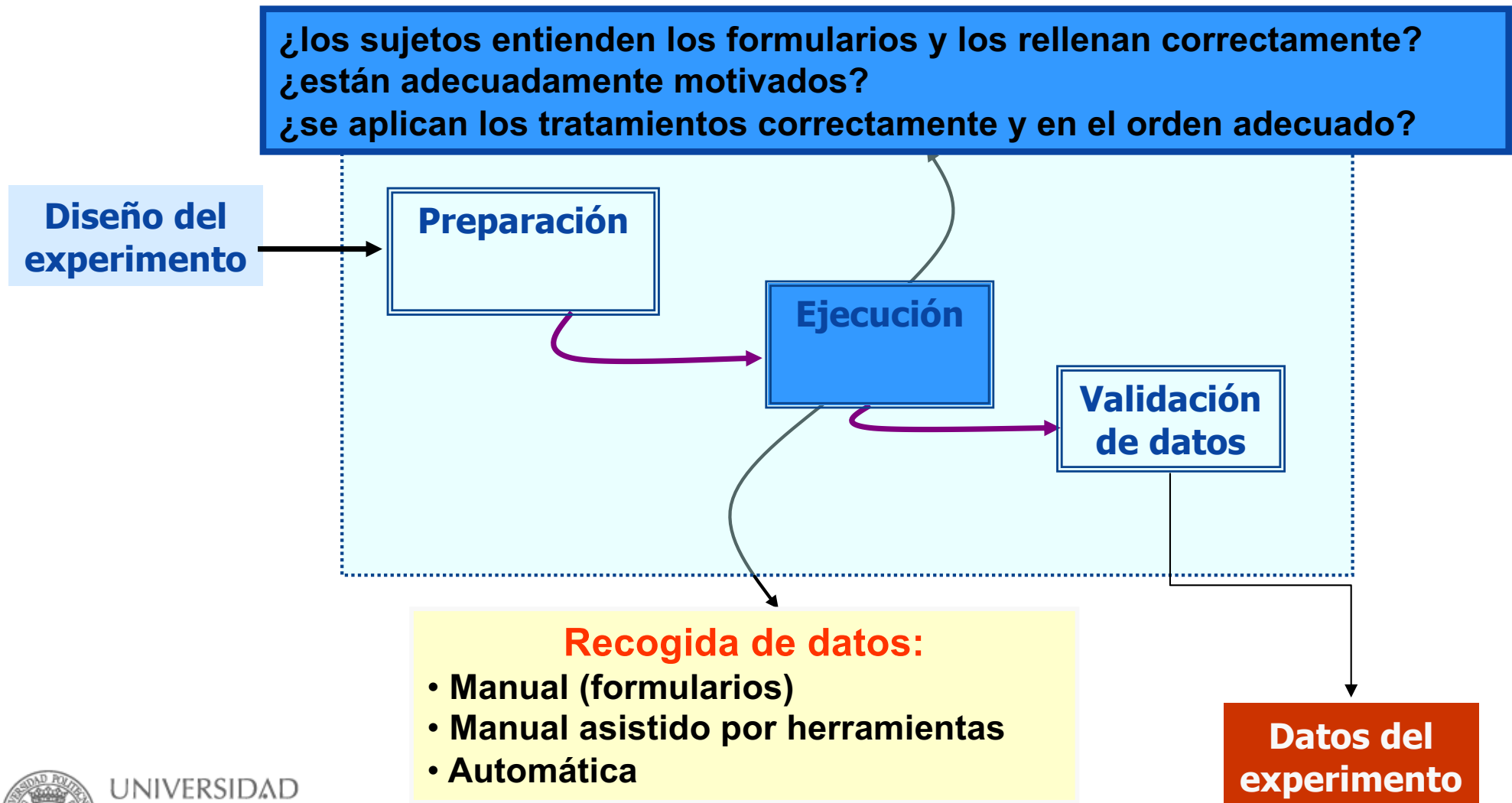
- Bajo poder estadístico, violación suposición tests
- Buscar o “pescar” un determinado resultado
- ...



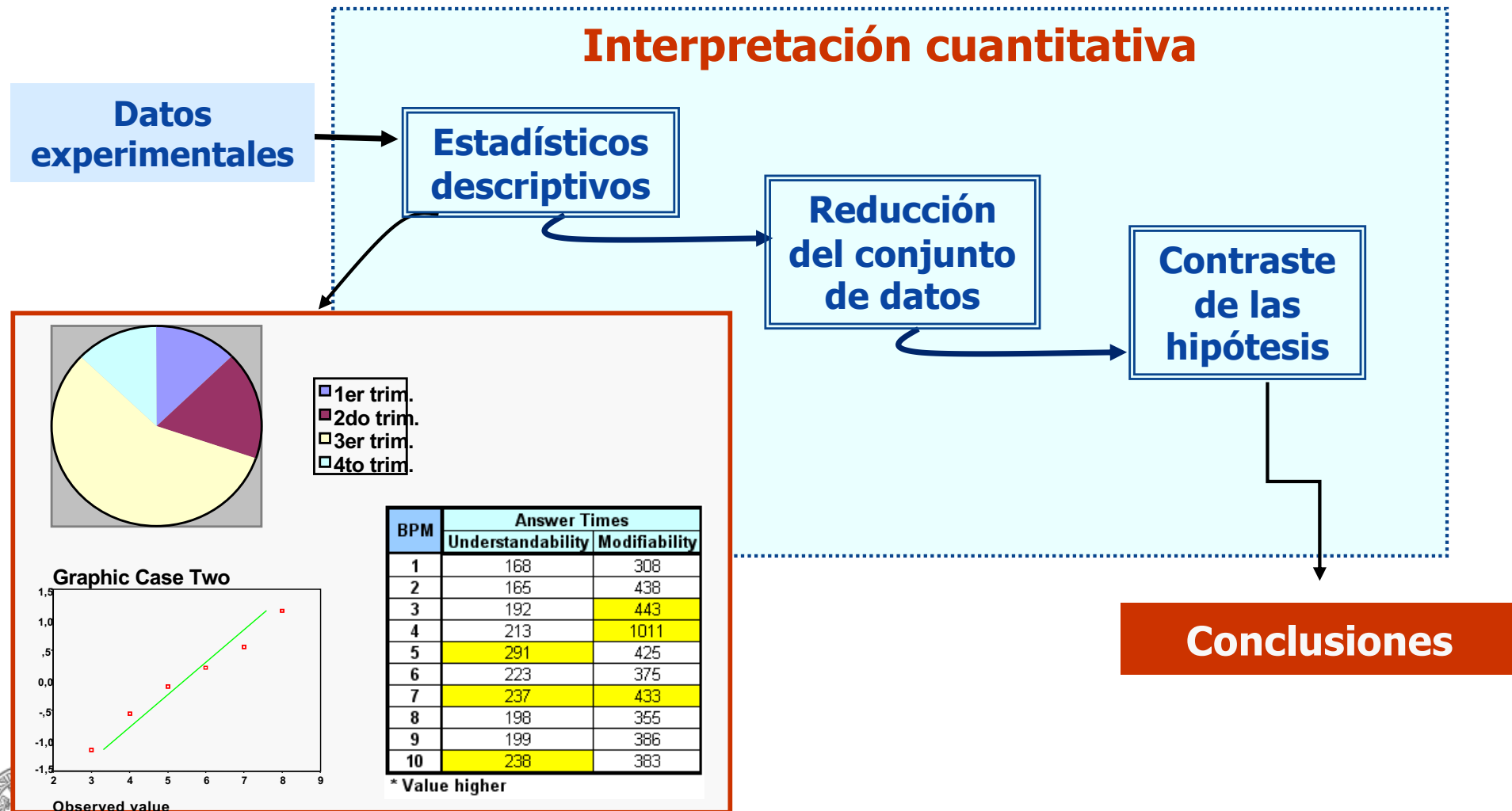
Proceso Experimental: Operación



Proceso Experimental: Operación



Proceso Experimental: Análisis e interpretación



Proceso Experimental: Análisis e interpretación

Contraste de Hipótesis



Diseño	Paramétrico	No paramétrico
Un factor, un tratamiento		Chi-2 Test binomial
Un factor, dos tratamientos	Test t Test F Test t emp	Mann-Whitney Chi-2 Wilcoxon Test de los signos
Un factor, más de dos tratamientos	ANOVA	Kruskal-Wallis Chi-2
Más de un factor	ANOVA	

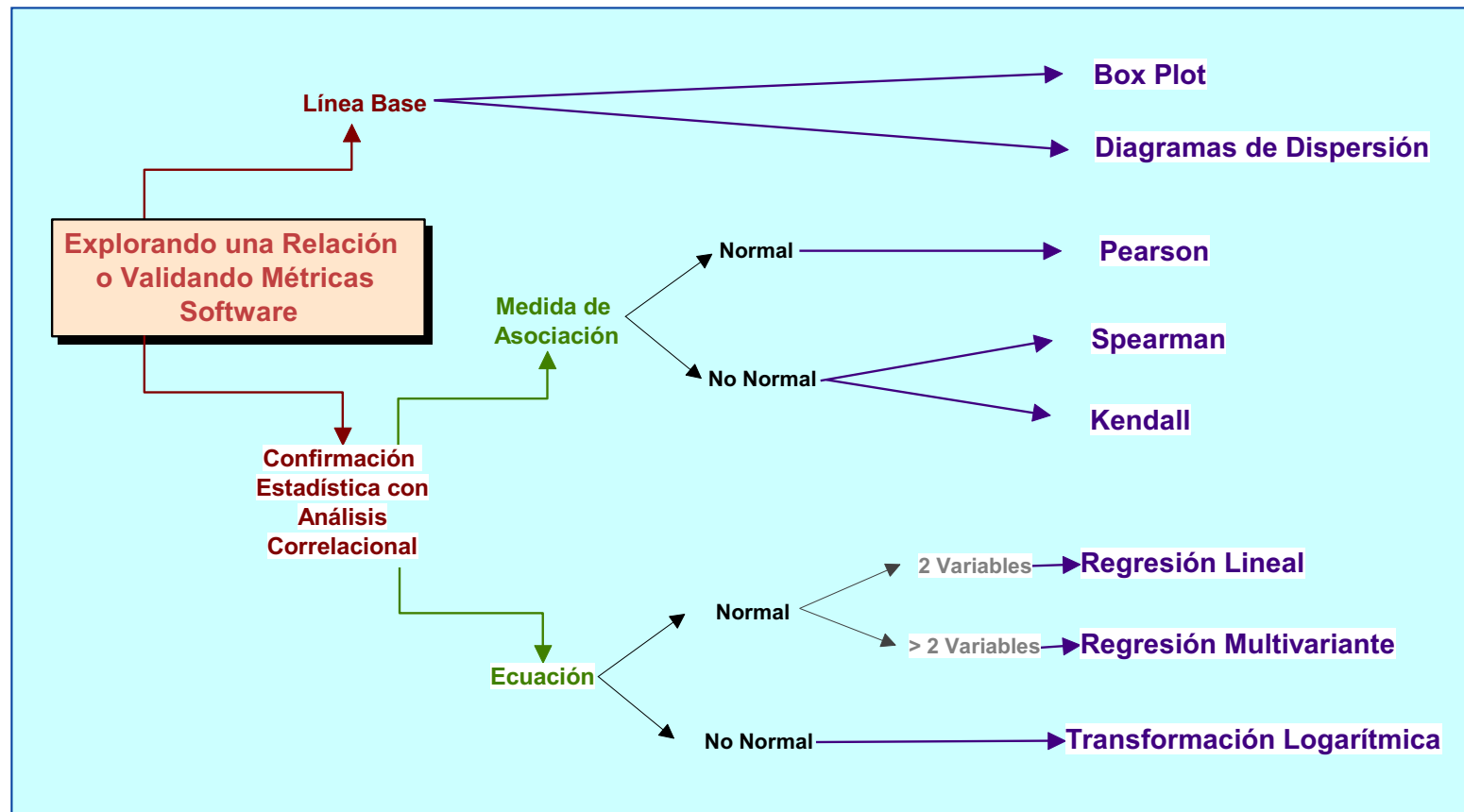
Otras Técnicas (IA) :

- Árboles de clasificación, Clustering, Redes neuronales, "Machine Learning"

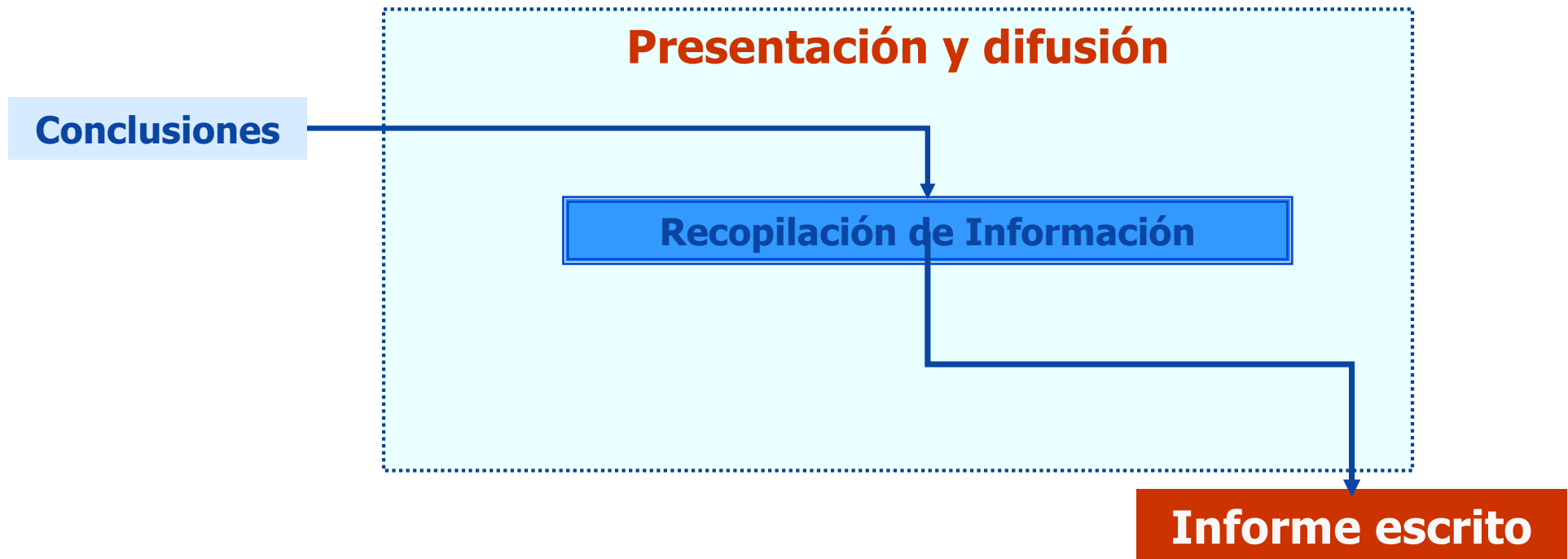


Proceso Experimental: Análisis e interpretación

Relación entre variables



Proceso Experimental: Presentación y difusión



Directrices para reportar experimentos
Jedlitschka et al. (2008)



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Proceso Experimental: Presentación y difusión



Sección	Sub-Sección	Alcance
Título		<título>+”-Un experimento controlado”
Autoría		¿Incluye información de contacto?
Resumen	Background	¿Por qué es importante esta investigación?
	Objetivo	¿Cuál es la pregunta tratada en esta investigación?
	Métodos	¿Cuál es el contexto estadístico y los métodos aplicados?
	Resultados	¿Cuáles son los resultados principales? Implicaciones prácticas?
	Limitaciones	¿Cuáles son las debilidades de esta investigación?
	Conclusiones	¿Cuál es la conclusión?
Palabras clave		Áreas de investigación, las variables dependientes, y el tipo del estudio
Introducción	Planteamiento del problema	¿Cuál es el problema? ¿Cuándo ocurre? ¿Quién lo ha observado? ¿Por qué es importante resolverlo?
	Objetivos de la investigación	Analizar <Objeto(s) de estudio> con el propósito de <propósito> con respecto a <foco de calidad> desde el punto de vista de <perspectiva> en el contexto de <contexto>
	Contexto	¿Cuáles son los factores ambientales que influyen en la generalización?

Directrices para reportar experimentos
Jedlitschka et al. (2008)



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Proceso Experimental: Presentación y difusión



Trabajo relacionado		¿Cómo se puede colocar esta investigación dentro de otra ya existente? Cómo se relaciona con la práctica?
Planificación del experimento	Objetivos	Formalización de los objetivos, definir los constructores importantes (e.g., el foco en la calidad) de los objetivos del experimento.
	Unidades experimentales	¿De qué población deberá ser la muestra mostrada? ¿Cómo se formarán los grupos (asignación de tratamientos)? Cualquiera que sea la clase de aleatorización debe ser descrita.
	Material Experimental	¿Qué objetos son seleccionados y por qué?
	Tareas	¿Qué tareas serán realizadas por los sujetos?
	Hipótesis y variables	¿Formalización de hipótesis y variables?
	Diseño	¿Qué tipo de diseño experimentas ha sido elegido?
	Procedimiento	¿Cómo se realizará (i.e., recolección de datos) el experimento? ¿Qué instrumentos, materiales, y herramientas se utilizarán y cómo?
	Procedimiento de Análisis	¿Cómo se analizarán los datos?
Ejecución		El propósito principal es describir cualquier desviación del plan. En caso de un número ilimitado de páginas la instanciación del plan (evitando redundancias) debe describirse, e.g., como se ve o describe la muestra (los participantes)?
	Preparación	¿Qué se ha hecho para preparar la ejecución del experimento (i.e, agenda, entrenamiento)
	Desviaciones	Describir cualquier desviación del plan, e.g, ¿Cómo fue realizada realmente la recolección de datos?



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Directrices para reportar experimentos
Jedlitschka et al. (2008)

Proceso Experimental: Presentación y difusión



Análisis	Estadísticas descriptivas	¿Cuáles son los resultados de la estadística descriptiva?
	Reducción del conjunto de datos	¿Fue necesario reducir el conjunto de datos, por qué y cómo?
	Contraste de las hipótesis	¿Cómo fueron evaluados los datos y fue validado el modelo de análisis?
Interpretación	Evaluación de los resultados e Implicaciones	Explicar los resultados y su relación con los resultados de la investigación anterior, especialmente aquellos mencionados en la sección de <i>Trabajo Relacionado</i>
	Amenazas a la validez	¿Cómo se asegura la validez de los resultados experimentales? ¿Cómo fueron validados los datos realmente? Amenazas que pudieran tener un impacto en los resultados (amenazas a la validez interna, variables confusas, prejuicios), y además, el grado en el cual la hipótesis captura los objetivos y la generalización de los resultados (amenazas a la validez externa, e.g, participantes, materiales) deben ser discutidos.
	Inferencias	Las inferencias deducidas de los datos a condiciones más generales
	Lecciones aprendidas	¿Qué experiencia fue recogida durante el curso del experimento?



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Directrices para reportar experimentos
Jedlitschka et al. (2008)

Proceso Experimental: Presentación y difusión



Conclusiones y trabajo futuro	Resumen	El propósito de esta sección es proveer un resumen conciso de la investigación y sus resultados según lo presentado en secciones anteriores.
	Impacto	Descripción de los impactos con respecto al coste, planificación, y calidad, circunstancias bajo las cuales la aproximación presumiblemente no producirá el beneficio esperado.
	Trabajo Futuro	¿Qué otros experimentos podrían emprenderse para investigar más profundamente los resultados obtenidos o para desarrollar el cuerpo del conocimiento?
Reconocimientos		Patrocinadores, participantes y colaboradores que no satisfagan los requisitos para la autoría deben ser mencionados
Referencias		Toda la literatura citada debe ser presentada en el formato propuesto por el editor
Apéndices		Materiales, datos originales, y análisis detallados, que pueden ser de ayuda para otros deben ser proporcionada

Directrices para reportar experimentos
Jedlitschka et al. (2008)



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Trabajo académico: Diseño de experimento (sesión 1)

- Consultar los **temas de trabajo** propuestos en PoliformaT
- Cada grupo deberá elegir un tema (entre los propuestos o proponer uno nuevo)
- Introducir el **tema e integrantes del grupo** aquí:
 - https://docs.google.com/document/d/15YxaDxUgbjZtQh9fC6pBdKjS9k_5_bxMjK-Zsj39l4s/edit?usp=sharing
- Definir los **objetivos del experimento** usando la plantilla del GQM (goal-question-metric)

Contenido

- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- **Ejemplo de experimento**
- Réplicas
- Agregación de resultados
- Recomendaciones
- Conclusiones
- Literatura relevante

Ejemplos de experimentos

- **Temas sobre experimentos relacionados con UML:**
 - Impacto de variantes en la notación sobre la comprensión y modificación
 - Impacto de la la calidad de los modelos UML en la comprensión y la modificación
 - Comparación de UML y ER sobre la comprensión y modificación
 - Validación de medidas como indicadores de la comprensión y modificación de los modelos UML
 - Beneficios de usar UML en el ciclo de vida del desarrollo de software





- **Assessing the influence of stereotypes on the comprehension of UML sequence diagrams: A controlled experiment**
 - Genero, M., Cruz-Lemus, J.A., Caivano, D., Abrahão, S., Insfran, E., Carsí, J.A.
 - 11th international conference on Model Driven Engineering Languages and Systems MoDELS'08 (LNCS 5301 pp. 280-294).



Material available at:
<http://alarcos.esi.uclm.es/ExpStereotypes>



UNIVERSIDAD
POLITECNICA
DE VALENCIA



■ Motivation

■ The influence of the use stereotypes in the comprehension of class diagrams has been studied

- F. Ricca, M. Di Penta, M. Torchiano, P. Tonella, M. Ceccato, **Developers' experience and ability influence web application comprehension tasks supported by UML stereotypes: a series of four experiments**, IEEE TSE 36 (1) (2010) 96-118.
- M. Staron, L. Kuzniarz, C. Wohlin, **Empirical assessment of using stereotypes to improve comprehension of UML models: a set of experiments**, The Journal of Systems and Software 79 (2006) 727-742

The influence of stereotypes on the comprehension of sequence diagrams has not been investigated yet





Motivation

- **UML sequence diagrams** → widely used technique for reasoning about object interactions
- **UML sequence diagram comprehension** → essential in the validation of requirements specifications among developers and stakeholders
- **A set of stereotypes for UML sequence diagrams** → proposed to enrich the semantics of interaction messages in the context of a Requirements Engineering approach for model-driven development (**signal, service, query, connect**)





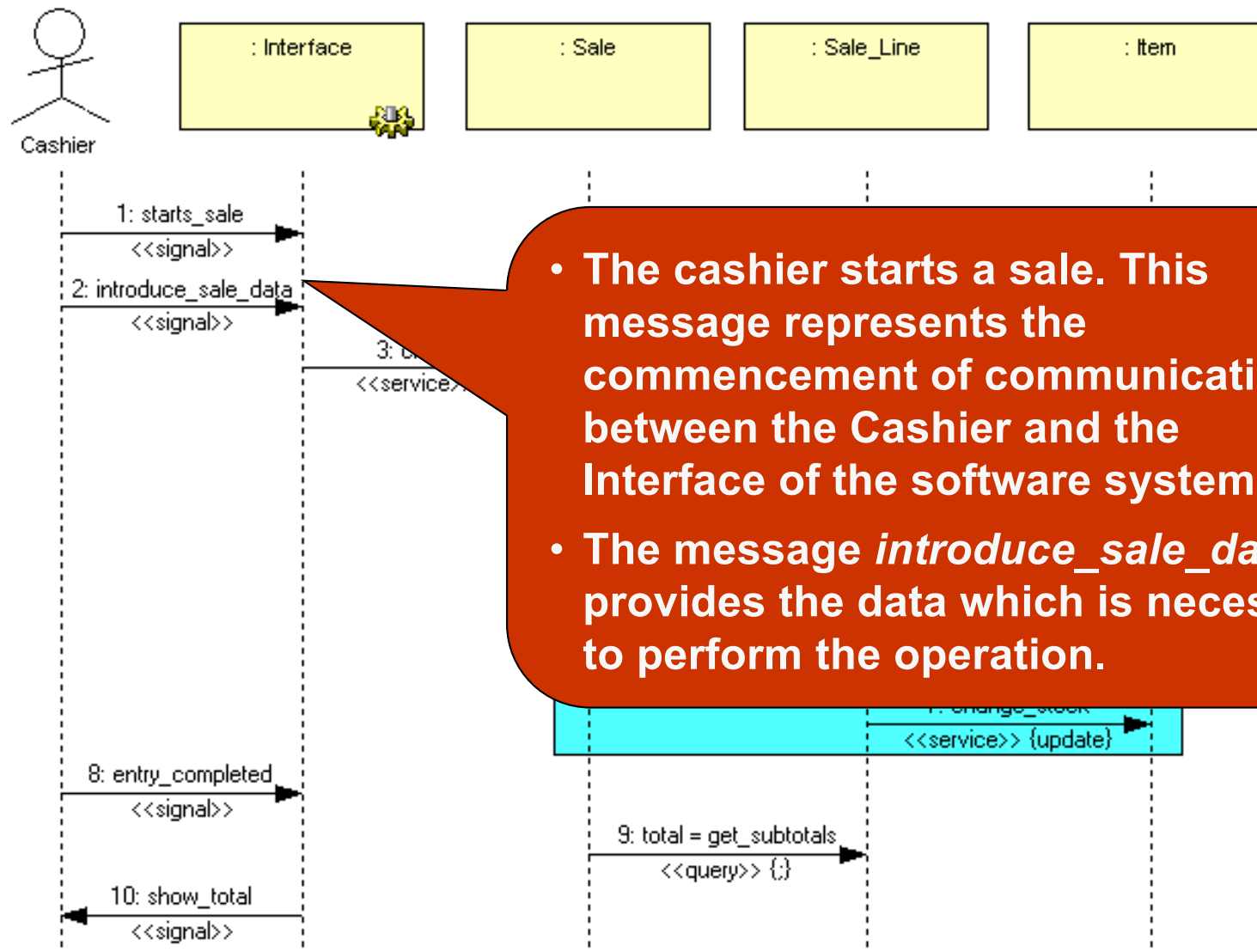
Stereotypes defined for Sequence Diagrams

- **«signal»** a message between an actor and the system.
- **«service»** a message that updates the state of an object. The properties *new/destroy/update* can be used if the object is to be created, destroyed or modified.
- **«query»** a message to query the state of a specific object or set of objects.
- **«connect»** a message which is used to indicate that an object from the sender object type needs to be connected (or related) to an object from the receiver object type.



Stereotypes for UML Sequence Diagrams

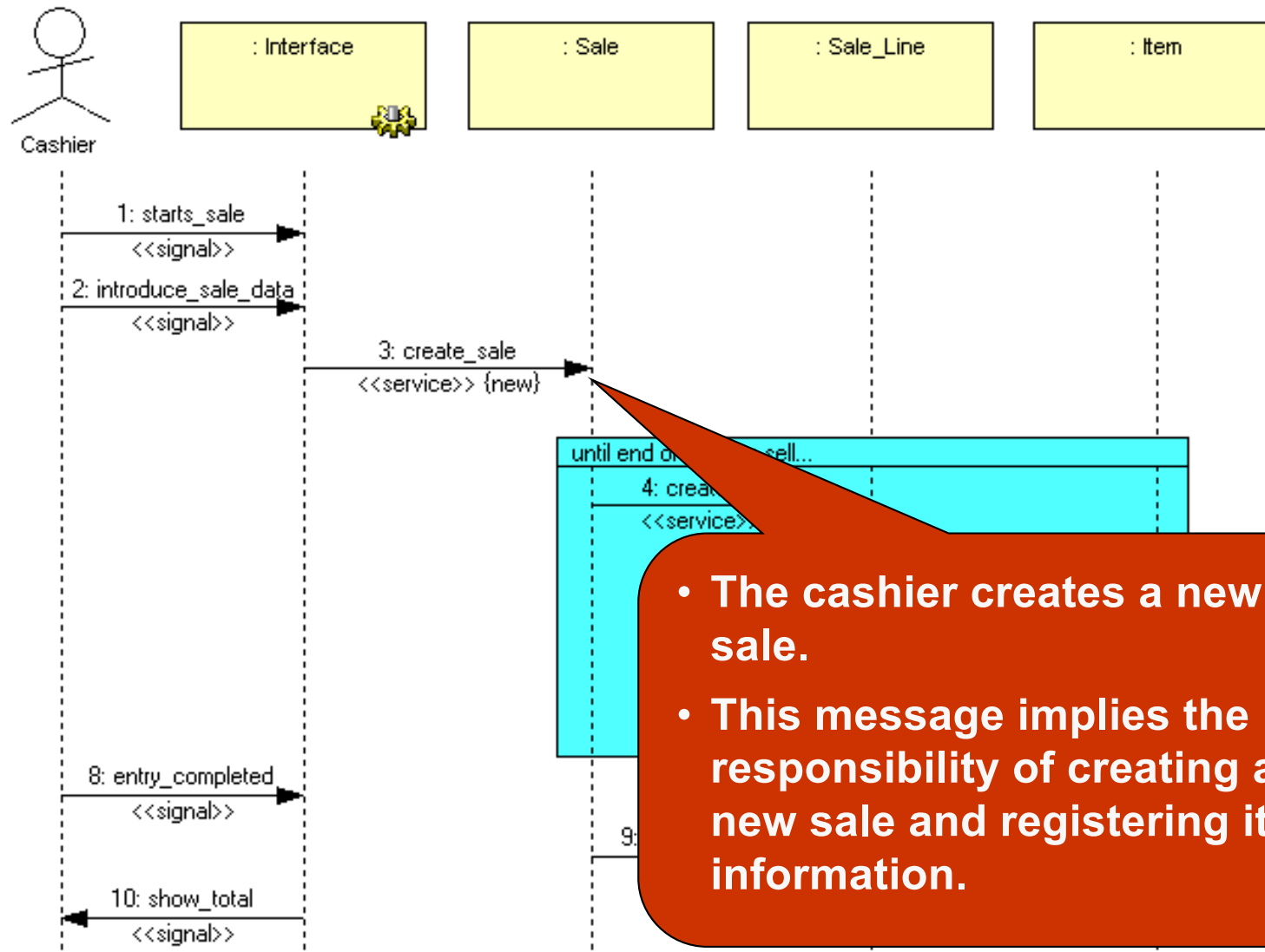
Items-Sale Use Case



- The cashier starts a sale. This message represents the commencement of communication between the Cashier and the Interface of the software system.
- The message *introduce_sale_data* provides the data which is necessary to perform the operation.

Stereotypes for UML Sequence Diagrams

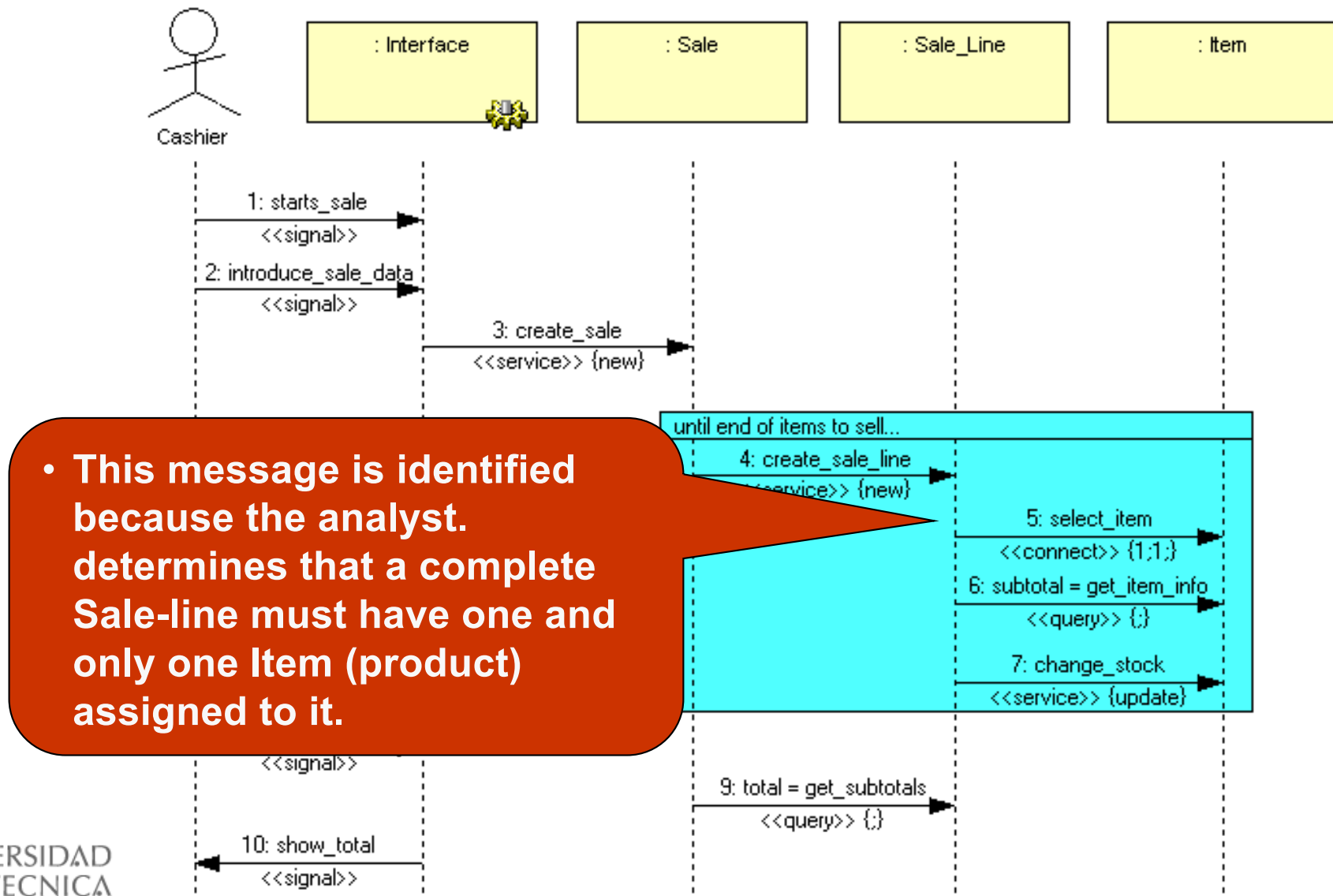
Items-Sale Use Case



- The cashier creates a new sale.
- This message implies the responsibility of creating a new sale and registering its information.

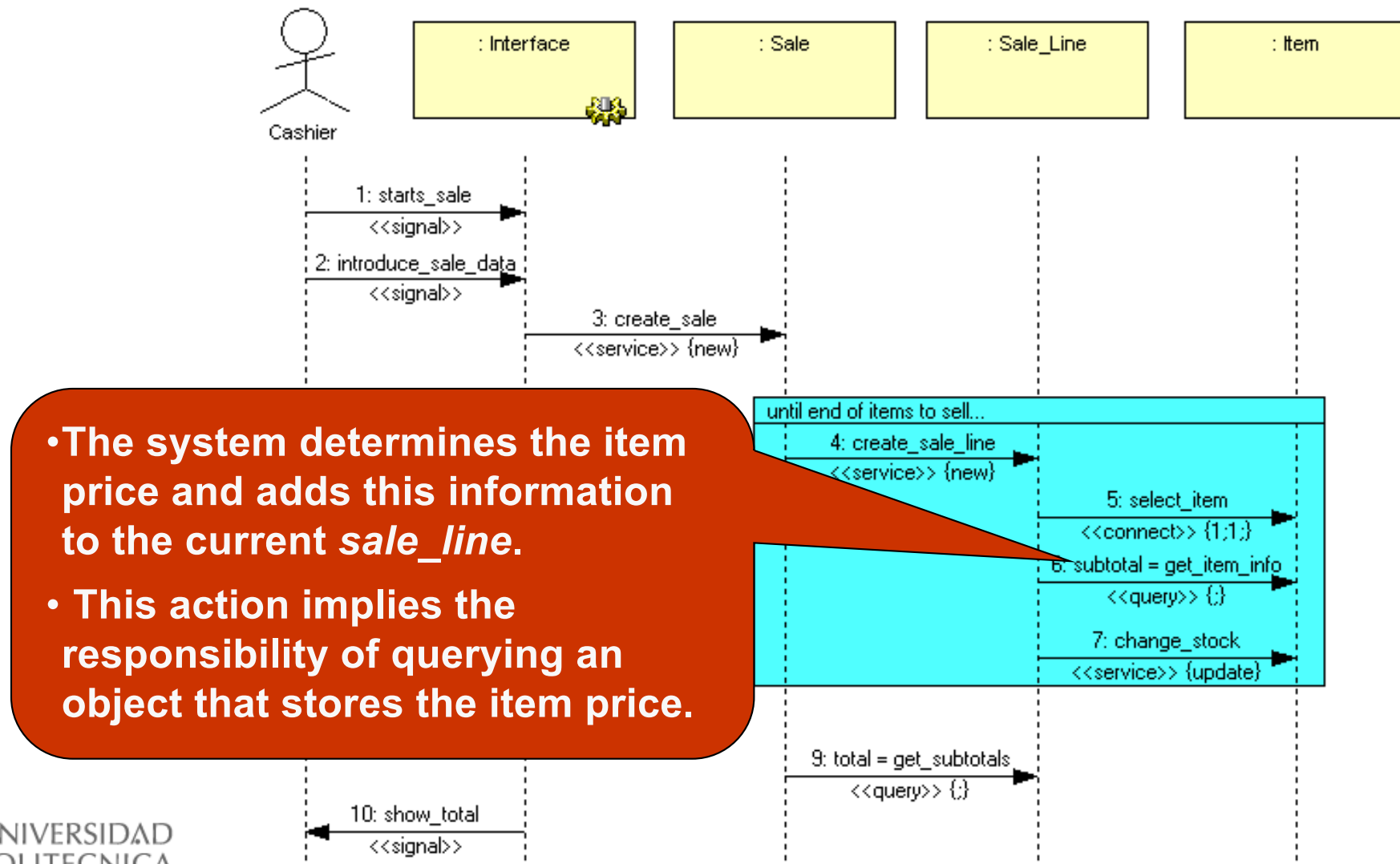
Stereotypes for UML Sequence Diagrams

Items-Sale Use Case



Stereotypes for UML Sequence Diagrams

Items-Sale Use Case





- We believe that the benefit of using these stereotypes is twofold :
 - Provide specific information about how to deal with each individual source element in the **automatic transformation of UML sequence diagrams into class diagrams**, in the MDD process.
 - **Improve the comprehension of UML sequence diagrams ???**

Goal → to assess whether the use of stereotypes improves the understanding of UML sequence diagrams





Independent variable(s)
(factors-treatment)

Diagram type

*(Stereotyped and
No-Stereotyped
SD)*

Effect

Dependent variable(s)
(outcome)

**Semantic
Comprehension**

Other independent variables
(controlled)
**(task complexity,
experience)**



UNIVERSIDAD
POLITECNICA
DE VALENCIA



Definition

- Using GQM template our goal is:
 - *Analyze the use of stereotypes*
 - *for the purpose of evaluating them*
 - *with respect to the comprehension of UML sequence diagrams*
 - *from the point of view of the researcher*
 - *in the context of fourth year undergraduate students in Computer Science from the University of Bari.*





- **Selection of subjects.** (students vs. professionals, probabilistic and non-probabilistic samples)
 - 77 fourth year Computer Science students
 - Selection by convenience

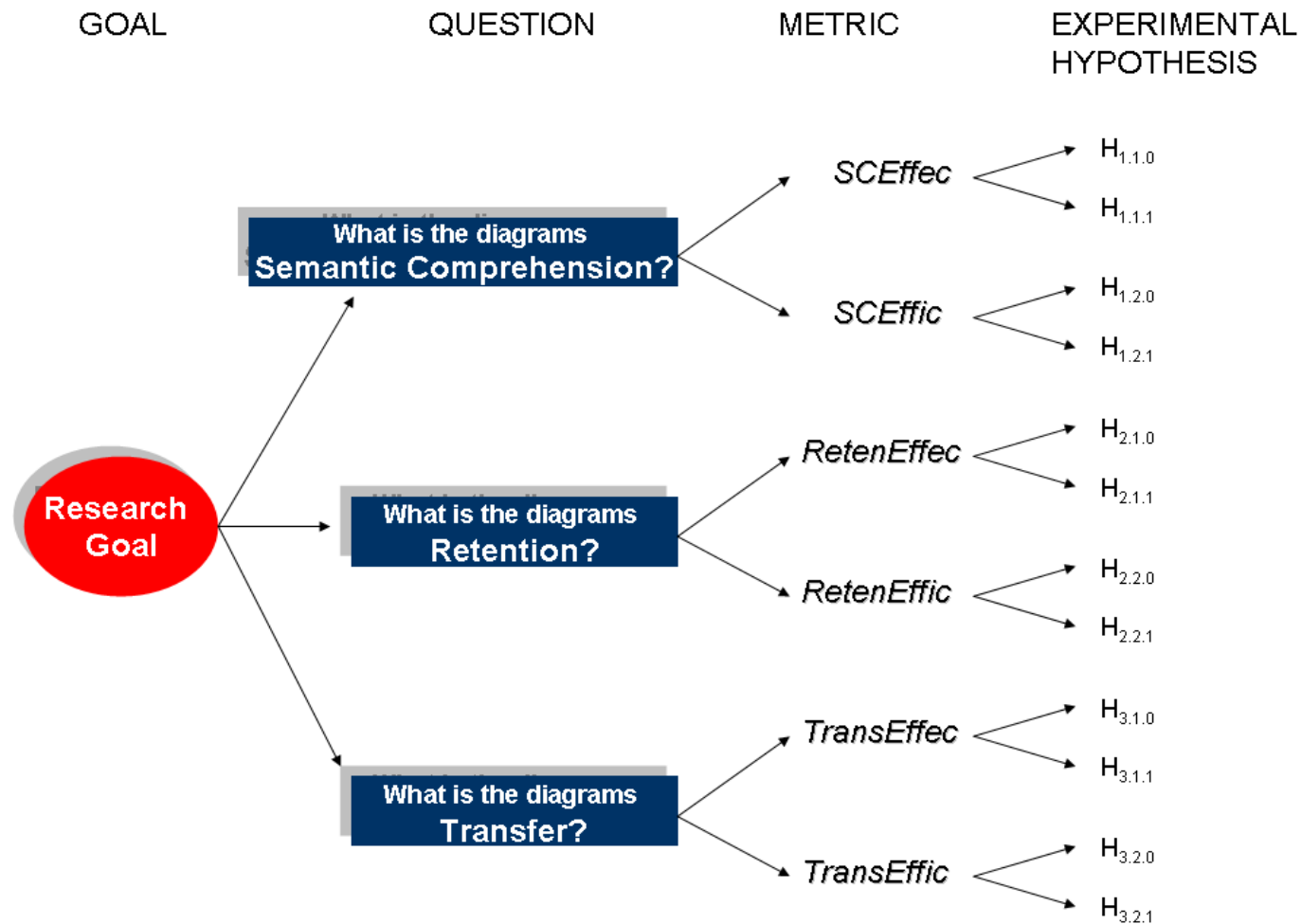
- **Selection of variables:**
 - **Independent (“factor”):**
 - *The diagram type*
 - Treatments: S (stereotyped) and N (non-stereotyped)

 - **Dependent (“outcome”, “effect”):**
 - *Semantic comprehension*: the ability to comprehend the semantics of the models.



- How to measure the dependent variable?
 - SCEffec: # correct answers / # questions
 - SCEffic: SCEffec / Time
- Hypotheses formulation
- Null Hypotheses (to be rejected):
 - *H1,0: stereotypes do not improve the subjects' SCEffec when attempting to comprehend a UML sequence diagram.*
 - *H2,0: stereotypes do not improve the subjects' SCEffic when attempting to comprehend an UML sequence diagram*
- Alternative Hypotheses:
 - *H1,1: stereotypes improve the subjects' SCEffec when attempting to comprehend a UML sequence diagram.*
 - *H2,1: stereotypes improve the subjects' SCEffic when attempting to comprehend an UML sequence diagram*

■ Variables and Experimental hypothesis





Relevant questions

- How many independent variables (factors)?
 - One factor → **Simple design**
 - More than one factor → **Factorial design**
- How many subjects per treatment ?
 - **Between subjects**: each subject receives only one treatment.
 - **Within subjects**: each subject receives all the treatments.

Independent variables → Diagram type + Domain
Domain (A-extras car rental, B-book hotel room)



RUN 1		DIAGRAM TYPE	
		S	N
DOMAIN	A	G1	G2
	B	G4	G3

RUN 2		DIAGRAM TYPE	
		S	N
DOMAIN	A	G3	G4
	B	G2	G1

A factorial within subjects design



- **Experiments objects can be obtained through:**
 - Make yourself
 - Make by students
 - Find in internet
 - Ask a company
 - Find in a repository (ReMoDD)

- **Experiment objects.** 4 diagrams corresponding to the four treatments prepared by the experimenters:
 - DA-S: stereotyped diagram, domain A (*extras car rental*)
 - DA-N: non-stereotyped diagram, domain A
 - DB-S: stereotyped diagram, domain B (*book hotel room*)
 - DB-N: non-stereotyped diagram, domain B





■ Experimental Tasks:

- Answer a test with 10 Yes/No questions concerning the semantics of the diagrams (SCEffec and SCEffic).
- Write down the start/end time.
- Manual tasks.



■ **Internal validity:** **Is it the treatment that really causes the effect?**

- Subjects with similar experiences
- Fatigue effect was avoided (time duration)
- Learning effect was avoided (material distributed in different order)
- Random assignment of subjects into groups

■ **External validity:** **Can the results be generalised?**

- Selection of objects and subjects

■ **Construct validity:** **Do the experimental settings actually reflect the construct under study?**

- Select well known domains
- Sequence diagrams with similar size
- Select appropriate measures
- Take care with evaluation apprehension

■ **Conclusion validity:** **Are the results statistically valid?**

- Select the appropriate statistical tests according the design
- Check assumptions of the statistical tests



Pilot study

- **Why?** Check the experimental material, the experiment duration.
- **When?** Two months before the experiment execution.
- **Who?** 4 PhD students of the Alarcos research group at UCLM (Spain).





Training session

- A training session of **three hours** took place the day before the experiment was carried out.
- The session included various tasks:
 - To **introduce** the students the **stereotypes** defined for UML sequence diagrams.
 - To **show** the students an **example**, similar to the material used in the experiment, in order to assure that the students had understood the experimental material and tasks.
 - We solved the tasks in the example with the students.





- The day after the training session, the experiment took place:
 - 30 minutes → explanation on how to perform the experiment
 - Randomly assignment of the subjects to the balanced groups
 - Experiment execution in two runs
- The experiment was conducted in a classroom, where the students were supervised and no communication among them was allowed.
- The groups were located in the same room.



Data analysis and interpretation (Descriptive statistics)

- The subjects obtained better global results in the stereotyped diagrams.
- To some extent, the **stereotyped diagrams** were **more comprehensible**, less error prone and required less time to understand than non-stereotyped diagrams.





Data analysis and interpretation (Hypothesis testing)

- Check normality of data
- Apply ANOVA test
- Results

Stereotypes

- There are no significant values.
- The power of the tests, i.e. the ability to detect an effect if there is one, are low.
- We cannot reject the null hypotheses investigated in this work and accepting the alternative hypotheses could be risky and misleading.

Domain

- In most cases, the domain affects the comprehension of the diagrams.
- The subjects performed worse with the diagrams of domain A (extras car rental). This could imply that the subjects found the extras car rental domain more difficult.





Data analysis and interpretation (Threats to validity)

After experiment execution threats must be analyzed

- **Internal validity:** We assumed a subjects' homogenous background, but this was not evaluated before the experiment was carried out.
- **Construct validity:** The extras car rental diagram was found to be much more difficult than that of the book hotel room.
- **External validity:** for generalising the results further replication is needed.



Conclusions

- Unfortunately the obtained results cannot be assumed to be final.
- The *p-values* observed in the hypotheses tests are **non-significant** and thus the **null-hypotheses cannot be rejected**.
- Nonetheless, descriptive statistics show a **slight tendency in favor of the use of stereotypes**.
- **Further replication is needed.**



Contenido

- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- Agregación de resultados
- Recomendaciones
- Conclusiones
- Literatura relevante

Réplicas

- “El uso de experimentos precisos y repetible es el signo de una disciplina madura o de ingeniería”
- Es importante replicar los experimentos (corroborar/refutar resultados):
 - **Internas:** réplicas realizadas por los investigadores que realizaron el experimento original.
 - **Externas:** réplicas realizadas por experimentadores que no tuvieron nada que ver con el experimento original.
- Es importante construir “paquetes de laboratorio”: descripción del proceso experimental, decisiones tomadas, etc.
 - Favorecen las réplicas externas
- Para lograr réplicas externas efectivas es necesario:
 - Comunicación efectiva entre el experimentador original y los experimentadores externos.
 - Transmitir el conocimiento tácito



Réplicas

- Las réplicas pueden ser:
 - Exactas: mismo protocolo de investigación y materiales. Es útil para comprender el rango de variabilidad de los resultados.
 - Diferenciadas: varia los elementos de la configuración del experimento base (original). Por ejemplo, variables, objetos experimentales.
- Para construir un cuerpo de conocimiento en SE, es necesario ejecutar **varios tipos de replicación**.
- Al realizar de manera sistemática diferentes tipos de replicación, los resultados experimentales se pueden verificar gradualmente



Contenido

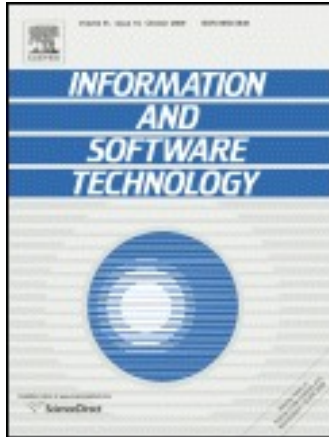
- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- **Agregación de resultados**
- Recomendaciones
- Conclusiones
- Literatura relevante

Agregación de resultados

- **Objetivo** → Agregar la evidencia empírica de experimentos que persiguen la misma hipótesis para construir un cuerpo de conocimiento sobre el tema investigado.
- **Técnicas para agregar resultados:**
 - **Mapeos y Revisiones sistemáticas de la literatura**
 - **Meta-análisis**
- **Beneficio del meta-análisis:**
 - **Agregación de datos** que conduce a un **poder estadístico más alto** y una estimación más robusta de lo que es posible obtener a partir del análisis de cualquier estudio individual.



Una familia de experimentos



- Cruz-Lemus, J.A., Genero, M., Caivano, D., Abrahão, S. Insfrán, E., Carsí, J.A.. **Assessing the influence of stereotypes on the comprehension of UML sequence diagrams: A family of experiments.** **Information and Software Technology** 53 (2011) 1391-1403
 - UBari: 77 students (Computer Science - 4th year) (Italy)
 - UCLM: 29 students (Computer Science - 3rd year) (Spain)
 - UCLM: 36 students (Computer Science - 5th year) (Spain)

The set of stereotypes presented in this work seem to be helpful for a better comprehension of UML sequence diagrams, especially with not well-known domains in the context of novice analysts.



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Contenido

- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- Agregación de resultados
- **Recomendaciones**
- Conclusiones
- Literatura relevante

Recomendaciones



- Kitchenham et al. (2002) proponen una serie de recomendaciones para la investigación empírica en la Ingeniería de Software.
- Jedlitschka et al. (2008) proponen recomendaciones sobre cómo reportar experimentos



Qué hacer?



Recomendaciones

■ Planificación

- Seleccionar objetos de tamaño y complejidad similar
- Asignar las mismas tareas a cada tratamiento
- Tratar de aliviar tanto como sea posible las amenazas a la validez
- Preparar pre-cuestionarios para recopilar la experiencia de los participantes
- Preparar post-cuestionarios para recopilar el feedback (opinion) de los participantes
- Preparar el material en el lenguaje nativo de los participantes



Qué hacer?



Recomendaciones

■ Operación

- Realizar un estudio piloto para chequear el material
- Hacer una estimación precisa del tiempo de duración
- Estimular a los participantes para que participen
- Capacitar a los participantes en el tema bajo estudio
- Antes de la ejecución del experimento realizar un ejemplo con los participantes
- Formar grupos balanceados por la experiencia
- Explicar claramente el procedimiento a seguir para la ejecución del experimento



Qué hacer?



Recomendaciones

■ **Análisis de datos e interpretación**

- Tener suficiente conocimiento estadístico o contactar a un experto
- Chequear las suposiciones de los test estadísticos
- Contextualizar los resultados



Qué hacer?



Recomendaciones

■ **Presentación y difusión**

- Seguir las directrices disponibles para reportar experimentos
- Reportar las amenazas que no se pudieron evitar
- Reportar la justificación de cada decisión tomada a lo largo del proceso experimental
- Preparar un “paquete de laboratorio” (material experimental, datos recolectados, etc.) y ponerlo disponible en la web.

■ **Realizar réplicas**



Qué NO hacer?



Recomendaciones

■ Definición



- No trate de validar “demasiadas cosas” en un único experimento

■ Planificación



- No desvele la hipótesis que se está investigando a los participantes antes de la realización del experimento
- No desvele cual es el tratamiento que estás tratando de probar que es mejor



Contenido

- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- Agregación de resultados
- Recomendaciones
- Conclusiones
- Literatura relevante

Conclusiones

- La **Ingeniería del Software Empírica** promueve la toma de decisiones en la Ingeniería del Software basándonos en la evidencia.
- La **Ingeniería del Software Empírica** proporciona los mecanismos para recopilar la evidencia empírica → **estudios empíricos**
- Los **experimentos** permiten evaluar metodos/técnicas en la práctica:
 - El feedback sobre las técnicas/métodos lleva a mejorar
- Es importante realizar una buena planificación del experimento
- El resultado de un experimento es un evento aislado. Es importante **replicar** los experimentos para confirmar/refutar sus resultados.
- Hay mucha bibliografía y ejemplos, tanto en contextos académicos como profesionales.



Contenido

- Experimentos
 - Conceptos generales
- Proceso experimental
 - Fases del proceso
- Ejemplo de experimento
- Réplicas
- Agregación de resultados
- Recomendaciones
- Conclusiones
- Literatura relevante

Literatura relevante

Directrices

- Kitchenham B., Pfleeger S., Pickard L., Jones P., Hoaglin D., El-Emam K. and Rosenberg J. (2002). **Preliminary Guidelines for Empirical Research in Software Engineering**. IEEE TSE, 28(8) 721-734.
- Jedlitschka, A., Ciolkowski, M. and Pfahl D. (2008). **Reporting experiments in Software Engineering**. Chapter 8, Guide to Advanced Empirical Software Engineering. Springer.
- Carver, J.. (2010). **Towards Reporting Guidelines for Experimental Replications: A Proposal**. 1st International Workshop on Replication in Empirical Software Engineering Research (RESER) [Held during ICSE 2010]. Cape Town, South Africa.
- Runeson, P., Höst, M. **Guidelines for conducting and reporting case study research in software engineering**. Empirical Software Engineering (2009) 14:131-164.
- Kitchenham, B. Pfleeger, S. **Principles of survey research**. ACM SIGSOFT Software Engineering Notes. Part 1 26(6) : 16-18 (2001), Part 2 27(1): 18-20 (2002), Part 3 27(2): 20-24 (2002), Part 4 27(3): 20-23 (2002), Part 5 27(5): 17-20 (2002), Part 6 28(2): 24-27 (2003).



Ejemplos de experimentos

- Dzidek, W. J., Arisholm, E., Briand, L. C. (2008). **A realistic empirical evaluation of the costs and benefits of UML in software maintenance.** IEEE Transactions on Software Engineering, 34(3), 407-432.
- Arisholm, E., Briand, L. C., Hove, S. E., & Labiche, Y. (2006). **The impact of UML documentation on software maintenance: An experimental evaluation.** IEEE Transactions on Software Engineering, 32(6), 365-381.
- Jeffrey C. Carver, Nachi Nagappan, Alan Page. **The Impact of Educational Background on the Effectiveness of Requirements Inspections: An Empirical Study.** IEEE Trans. Software Eng. 34(6): 800-812 (2008)
- Erik Arisholm, Hans Gallis, Tore Dybå, Dag I. K. Sjøberg. **Evaluating Pair Programming with Respect to System Complexity and Programmer Expertise.** IEEE Trans. Software Eng. 33(2): 65-86 (2007)
- Lionel C. Briand, Yvan Labiche, Massimiliano Di Penta, Han (Daphne) Yan-Bondoc. **An Experimental Investigation of Formality in UML-Based Development.** IEEE Trans. Software Eng. 31(10): 833-849 (2005)



Literatura relevante

Ejemplos de experimentos

- Silvia Teresita Acuña, Marta Gómez, Natalia Juristo Juzgado. **How do personality, team processes and task characteristics relate to job satisfaction and software quality?** Information & Software Technology 51(3): 627-639 (2009)
- Scaniello, G., Gravino, C., Genero, M., Cruz-Lemus, J.A., Tortora, G. **The Impact of UML Models Produced in the Requirements Analysis Process on Source Code Comprehensibility and Modifiability: Results from a Family of Four Experiments.** ACM TOSEM (2013).
- Carmine Gravino, Emilio Insfrán, Giuseppe Scanniello, Genoveffa Tortora: **Assessing the Effectiveness of Sequence Diagrams in the Comprehension of Functional Requirements: Results from a Family of Five Experiments.** IEEE Trans. Software Eng. 39(3): 327-342 (2013)
- Julio Sandobalin, Emilio Insfrán, Silvia Abrahão: **On the Effectiveness of Tools to Support Infrastructure as Code: Model-Driven Versus Code-Centric.** IEEE Access 8: 17734-17761 (2020)



- Basili V., Shull F. and Lanubile F. (1999). **Building knowledge through families of experiments**, IEEE TSE, 25(4) 435-437.
- Brooks, A., Roper, M., Wood, M., Daly, J., Miller, J. (2008). **Replication's role in Software Engineering**. Chapter 14, Guide to Advanced Empirical Software Engineering. Springer.
- Shull, F., Carver, J., Vegas, S., Juristo, N. (2008). **The role of replications in Empirical Software Engineering**. Empirical Software Engineering 13(2), 211-218.
- Santos A., Vegas S., Oivo M., Juristo N.: **Comparing the results of replications in software engineering**. Empir. Softw. Eng. 26(1): 13 (2021)
- Santos A., Vegas S., Oivo M., Juristo N.: **A Procedure and Guidelines for Analyzing Groups of Software Engineering Replications**. IEEE Trans. Software Eng. 47(9): 1742-1763 (2021)

