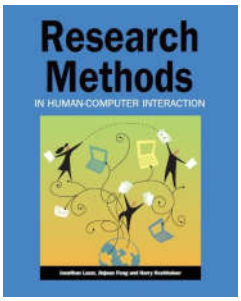
BIRZEIT UNIVERSITY

Chapter 2

Experimental Research



By: Mamoun Nawahdah (PhD)
2018

Overview

- Types of behavioral research
- Research **hypotheses**
- Basics of experimental research
- Significance tests
- Limitations of experimental research

Types of Behavioral Research

- **Descriptive investigations:** focus on constructing an accurate description of what is happening.
- **Relational investigations:** enable the researcher to identify relations between multiple factors. However, relational studies can rarely determine the causal relationship between multiple factors.
- **Experimental research:** allows the establishment of causal relationship.

Types of Behavioral Research

Type of research	Focus	General claims	Typical methods
Descriptive	Describe a situation or a set of events	X is happening	Observations, field studies, focus groups, interviews
Relational	Identify relations between multiple variables	X is related to Y	Observations, field studies, surveys
Experimental	Identify causes of a situation or a set of events	X is responsible for Y	Controlled experiments

Relationship between descriptive research, relational research, and experimental research.

Research Hypotheses

- An experiment normally starts with a research hypothesis.
- A hypothesis is a ***precise problem statement that can be directly tested through an empirical investigation***.
- Compared with a **theory**, a hypothesis is a **smaller**, **more focused** statement that can be examined by a single experiment.

Types of Hypotheses

- **Null hypothesis**: typically states that there is **NO** difference between experimental treatments.
- **Alternative hypothesis**: a statement that is **mutually exclusive** with the null hypothesis.
- The goal of an experiment is to find **statistical evidence** to disprove or nullify the null hypothesis in order to support the alternative hypothesis.
- A hypothesis should specify the **independent variable(s) (IV)** and **dependent variable(s) (DV)**.

Research Hypotheses

- **Independent variables (IV)** refer to the factors that the researchers are interested in studying or the possible "**cause**" of the change in the dependent variable.
  - **IV** is independent of a participant's behavior.
  - **IV** is usually the treatments or conditions that the researchers can control.
- **Dependent variables (DV)** refer to the outcome or effect that the researchers are interested in.
  - **DV** is dependent on a participant's behavior or the changes in the **IV**s.
  - **DV** is usually the outcomes that the researchers need to measure.

Typical IV in HCI

- Those that relate to technology:
  - Types of technology or device.
  - Types of design.
- Those that relate to users: age, gender, computer experience, professional domain, education, culture, motivation, mood, and disabilities.
- Those that relate to context of use:
  - Physical status.
  - User status.
  - Social status.

Typical DV in HCI

- **Efficiency**:
  - e.g., task completion time, speed
- **Accuracy**:
  - e.g., error rate
- **Subjective satisfaction**:
  - e.g., Likert scale ratings
- **Ease of learning and retention rate**
- **Physical or cognitive demand**
  - e.g., NASA task load index

Components of Experiment

- **Treatments, or conditions**: the different techniques, devices, or procedures that we want to compare.
- **Subjects/Units**: the objects to which we apply the experiment treatments.
  - In HCI research, the subjects/units are normally human subjects with specific characteristics, such as gender, age, or computing experience.
- **Assignment method**: the way in which the experimental subjects/units are assigned different treatments.

Randomization

- **Randomization**: the random assignment of treatments to the experimental units or participants.
- In a totally randomized experiment, no one, including the investigators themselves, is able to predict the condition to which a participant is going to be assigned.

Significance Tests

- **Why do we need significance tests?**
  - When the values of the members of the comparison groups are **all** known, you can directly compare them and draw a conclusion. No significance test is needed since there is no **uncertainty** involved.
  - When the population is large, we can only sample a **sub-group** of people from the entire population.
  - Significance tests allow us to determine how **confident** we are that the results observed from the sampling population can be **generalized** to the entire population.

Type I and Type II Errors

- All significance tests are subject to the risk of **Type I** and **Type II** errors:
  - A **Type I** error (also called an ***α*** *error or a "**false positive**") refers to the mistake of* **rejecting the null hypothesis** when it is true and should not be rejected.
  - A **Type II** error (also called a ***β*** *error or a "**false negative**") refers to the mistake of **not rejecting the null*** **hypothesis** when it is false and should be rejected.

Type I and Type II Errors

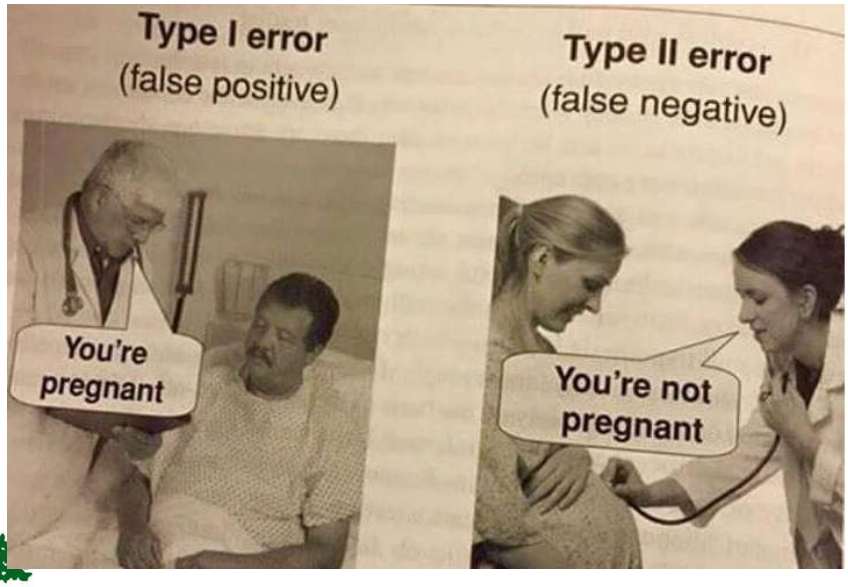
		Jury decision	
		Not guilty	Guilty
Reality	Not guilty	✓	Type I error
	Guilty	Type II error	✓

Type I and Type II errors in the judicial case.

		Study conclusion	
		No difference	Touchscreen ATM is easier to use
Reality	No difference	✓	Type I error
	Touchscreen ATM is easier to use	Type II error	✓

Type I and Type II errors in a hypothetical HCI experiment.

Type I and Type II Errors



Type I and Type II Errors

- It is **generally** believed that **Type I** errors are worse than **Type II** errors.
- Statisticians call **Type I** errors a mistake that involves "naivety سذاجة".
  - A **Type I** error may result in a condition worse than the current state.
- **Type II** errors are mistakes that involve "blindness عمى"
  - A **Type II** error can cost the opportunity to improve the current state.

Example 1

- Hypothesis: "Adding water to toothpaste protects against cavities."
  - Null hypothesis (H0): "Adding water to toothpaste has no effect on cavities."
  - A type I error occurs when detecting an effect (adding water to toothpaste protects against cavities) that is not present.
  - The null hypothesis is true (i.e., it is true that adding water to toothpaste has no effect on cavities), but this null hypothesis is rejected based on bad experimental data.

Example 2

- Hypothesis: "Adding fluoride to toothpaste protects against cavities."
  - Null hypothesis (H0): "Adding fluoride to toothpaste has no effect on cavities."
  - A type II error occurs when failing to detect an effect (adding fluoride to toothpaste protects against cavities) that is present.
  - The null hypothesis is false (i.e., adding fluoride is actually effective against cavities), but the experimental data is such that the null hypothesis cannot be rejected.

Example 3

- Hypothesis: "The evidence produced before the court proves that this man is guilty.“
  - Null hypothesis (H0): "This man is innocent.“
  - A type I error occurs when convicting an innocent person (a miscarriage of justice).
  - A type II error occurs when letting a guilty person go free (an error of impunity).
  - A positive correct outcome occurs when convicting a guilty person. A negative correct outcome occurs when letting an innocent person go free.

Example 4

- Hypothesis: "A patient's symptoms improve after treatment A more rapidly than after a placebo treatment.“
  - Null hypothesis (H0): "A patient's symptoms after treatment A are indistinguishable from a placebo.“
  - A Type I error would falsely indicate that treatment A is more effective than the placebo, whereas a
  - Type II error would be a failure to demonstrate that treatment A is more effective than placebo even though it actually is more effective.

Controlling Risks of Errors

- In statistics, the probability of making a **Type I** error is called ***alpha*** (or **significance level**, **p value**).
- The probability of making a **Type II** error is called ***beta***.
- The **statistical power** of a test, defined as **1–*β***, refers to *the probability of successfully rejecting a **null hypothesis*** when it is false and should be rejected.

Controlling Risks of Errors

- ***Alpha*** and ***Beta*** are interrelated. Under the same conditions, decreasing alpha reduces the chance of making **Type I** errors but increases the chance of making **Type II** errors.
- In experimental research, it is generally believed that **Type I** errors are worse than **Type II** errors.
- So a very low **p** value (**0.05**) is widely adopted to control the occurrence of **Type I** errors.

Limitations of Experimental Research

- Experimental research requires well-defined, testable hypotheses that consist of a limited number of dependent and independent variables.
- Experimental research requires strict control of factors that may influence the dependent variables.
- Lab-based experiments may not be a good representation of users' typical interaction behavior.