

Scientific Methodology in Computer Science

MO430A

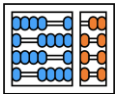
Prof. Dr. Bruno B. P. Cafeo

Institute of Computing
University of Campinas

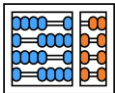
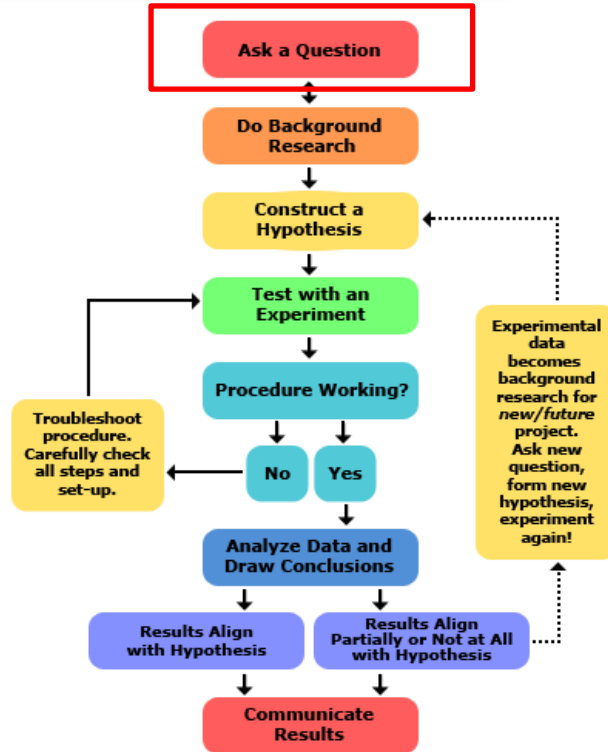


Agenda

- Formulating research questions
- Developing hypotheses
- The role of hypotheses in research
- Examples of research questions in computer science

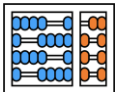


Recap



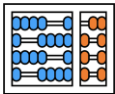
Introduction

- In research, asking clear and answerable research questions (RQs) is important
- The data (evidence) that must be collected depends on the RQ
- In quantitative research, summarising and analysing the data typically uses numerical methods (such as averages or percentages), so the RQs must be appropriate for analysis using these methods.
- RQs emerge from observations, which leads to asking questions, and the need for evidence to answer that question



Introduction

- The RQ drives all other aspects of the research



Introduction

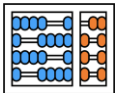
- Defining the RQ precisely can be challenging
- Studies often have an overall, broad research goal with many sub-question (which may be quantitative or qualitative)
- *How can a microservice architecture support efficiently enable architecting and implementing data science workflows?*
 - *(SRQ1) How can the scope of services regarding a microservice architecture be identified?*
 - *(SRQ2) To what extent can functional and non-functional requirements for data science workflows be formulated?*
 - *(SRQ3) How does a microservice architecture for data science workflows have to be determined to fulfil the identified functional and non-functional requirements from SRQ2?*



Schröer C. Towards microservice identification approaches for architecting data science workflows. *Procedia Computer Science*, 2021, 181: 519–525

Elements of RQs

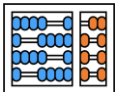
- A RQ must be written carefully so it can be properly answered
- RQs usually present four potential components:
 - Population
 - Intervention
 - Comparison
 - Outcome



The Population

The *population* is the group of *individuals* (or *cases*; or *subjects* if the individuals are people) from which the total set of observations of interest could be made, and to which the results will (hopefully) generalise.

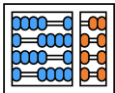
- The population represents all the ‘individuals’ to which the results are to be generalized.
 - Individuals or cases do not just refer to people., though the words may be commonly used that way.
 - Similarly, population does not just mean people. In this context, a population is any group of interest.
- In contrast, the sample is the subset of the population that we actually end up studying, from which data are obtained.



The Intervention

An intervention is a comparison or connection that the researchers have imposed upon those in the study, intending to change the outcome.

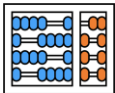
- It is usually seen as a possible solution
- Application of a theory or method



The Outcome

The outcome in a RQ is the result, output, consequence or effect of interest in a study, numerically summarising the population (or subsets of the population).

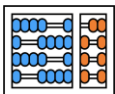
- Because the RQ concerns a population, the outcome describes a population as a whole.
- The outcome is usually an average, percentage, or general quantity numerically summarising the population (or subsets of the population)



The Comparison

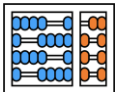
The comparison in the RQ identifies the small number of different, distinct subsets of the population between which the outcome is being compared. The groups being compared have either imposed differences or have existing differences.

- The population represents all the ‘individuals’ to which the results are to be generalized.
 - Individuals or cases do not just refer to people., though the words may be commonly used that way.
 - Similarly, population does not just mean people. In this context, a population is any group of interest.
- In contrast, the sample is the subset of the population that we actually end up studying, from which data are obtained.



Types of RQs

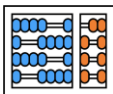
- All RQs have a population (P) and an outcome (O)
- However, different types of RQ emerge depending on whether the RQ also has an comparison/connection (C) or intervention (I)
 - Descriptive RQs (PO)
 - Relational RQs (PCO)
 - Interventional RQs (PICO)



Descriptive RQs (PO)

- Descriptive RQs are the most basic RQs and identify:
 - The **P**opulation to be studied
 - The **O**utcome of interest about this population
- Typically, descriptive RQs look like this:
 - Among {the population}, what is {the outcome}?

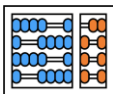
What is the DevOps adoption rate within small-sized businesses?



Relational RQs (PCO)

- Usually, relationships are more interesting than just descriptions; relational RQs explore existing relationships. Relational RQs identify:
 - The **P**opulation to be studied
 - The **C**omparison (or Connection)
 - The **O**utcome of interest about this population
- Relational RQs have no intervention; the connection or comparison is not imposed by the researchers.
- Typically, relational RQs look like this:
 - Among {the population}, is {the outcome} the same for {the groups being compared}?

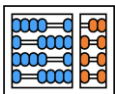
How does the proportion of DevOps adoption differ among small-sized companies compared to medium-sized companies?



Interventional RQs (PICO)

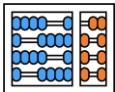
- Interventional RQs explore relationships where the comparison/connection is determined or allocated by the researchers. They identify:
 - The **P**opulation to be studied
 - The **I**ntervention
 - The **C**omparison (or Connection)
 - The **O**utcome of interest about this population
- Interventional RQs may look like relational RQs, except that the comparison or connection is determined or allocated (i.e., imposed) by the researchers.
- When writing interventional RQs, make efforts to make it clear, if possible, when the RQ is interventional

How does the implementation of DevOps practices (I) affect the proportion of DevOps adoption (O) among small-sized companies (P) compared to their counterparts in medium-sized companies (C)?



Two approaches to RQ

- RQs can be approached in one of two ways:
 - **For estimation (confidence intervals):** These RQs are concerned with, for example, estimating a value in a population. This value may be the size of a difference (probably a RQ with a Comparison), or strength of a relationship (probably a RQ with a Connection).
 - **For making decisions (hypothesis testing):** These RQs are concerned with making a decision about an unknown population value: for example, is the percentage the same in two different groups of the population?

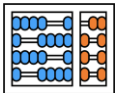


Writing good RQs

- Ideally, a well-written should be (FINER):
 - **Feasible:** Answering the RQ should be possible practically; sufficient personnel, time, resources, and money should be available to complete the study properly.
 - **Interesting:** The RQ should be interesting. For example, no-one cares about comparing the percentage of people who prefer drinking tea in blue cups to green cups...
 - **Novel:** The RQ should be original (the RQ should 'seek to confirm, refute or extend previous findings, and potentially reveal new findings'. Researching something already well known is waste of time and resources.
 - **Ethical:** The RQ must be able to be answered ethically. This is not negotiable.
 - **Relevant:** The RQ should be relevant and current.

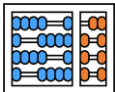


Anastasiadis E, Rajan P, Winchester CL. Framing a research question: The first and most vital step in planning research. *Journal of Clinical Urology*. 2015;8(6):409–11.



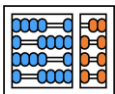
Conceptual and operational definitions

- Research studies usually include terms that must be carefully and precisely defined, so that others know exactly what has been done and there are no ambiguities.
- **Conceptual definition:** A conceptual definition articulates what exactly is to be measured or observed in a study.
- **Operational definition:** An operational definition articulates how to capture (identify, create, measure, assess etc.) the value.



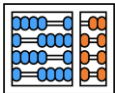
Estimation: Confidence intervals questions

- The RQ concerns how precisely a value in the population is estimated by the sample.
- This value may measure a difference, or the strength of a relationship.



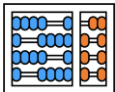
Making decisions: Hypothesis testing questions

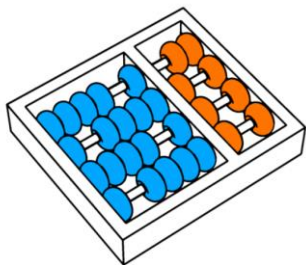
- Sometimes, RQs are not about the precision with which a population value is estimated by the sample, but instead about deciding if a difference or a relationship exists in the population.
- These RQs often are associated with hypotheses: statements that suggest possible answers to the RQ.



Homework

- Read Articles:
 - Select and read one or two research articles in your field.
- Identify Research Questions (RQs):
 - For each article, identify the main research question (RQ). What is the central problem or issue the study is investigating?
- Identify Sub Research Questions (Sub RQs):
 - For each RQ, identify any sub-research questions or secondary questions that are used to support the main RQ.
- Assess PICO Components:
 - Break down the RQs into PICO components (Population, Intervention, Comparison, Outcome). Identify if they are explicitly stated.
- Determine RQ Type:
 - Categorize each RQ as descriptive, relational or interventional
- Evaluate Operational and Conceptual Definitions:
 - For each RQ and sub-RQ, evaluate the operational and conceptual definitions used. How are key terms and variables defined in the article?





**INSTITUTO DE
COMPUTAÇÃO**



Prof. Dr. Bruno B. P. Cafeo

Sala 04
Instituto de Computação - Unicamp
Av. Albert Einstein, 1251
Cidade Universitária
Campinas – SP
13083-852

<https://ic.unicamp.br/~cafeo/>
cafeo@ic.unicamp.br