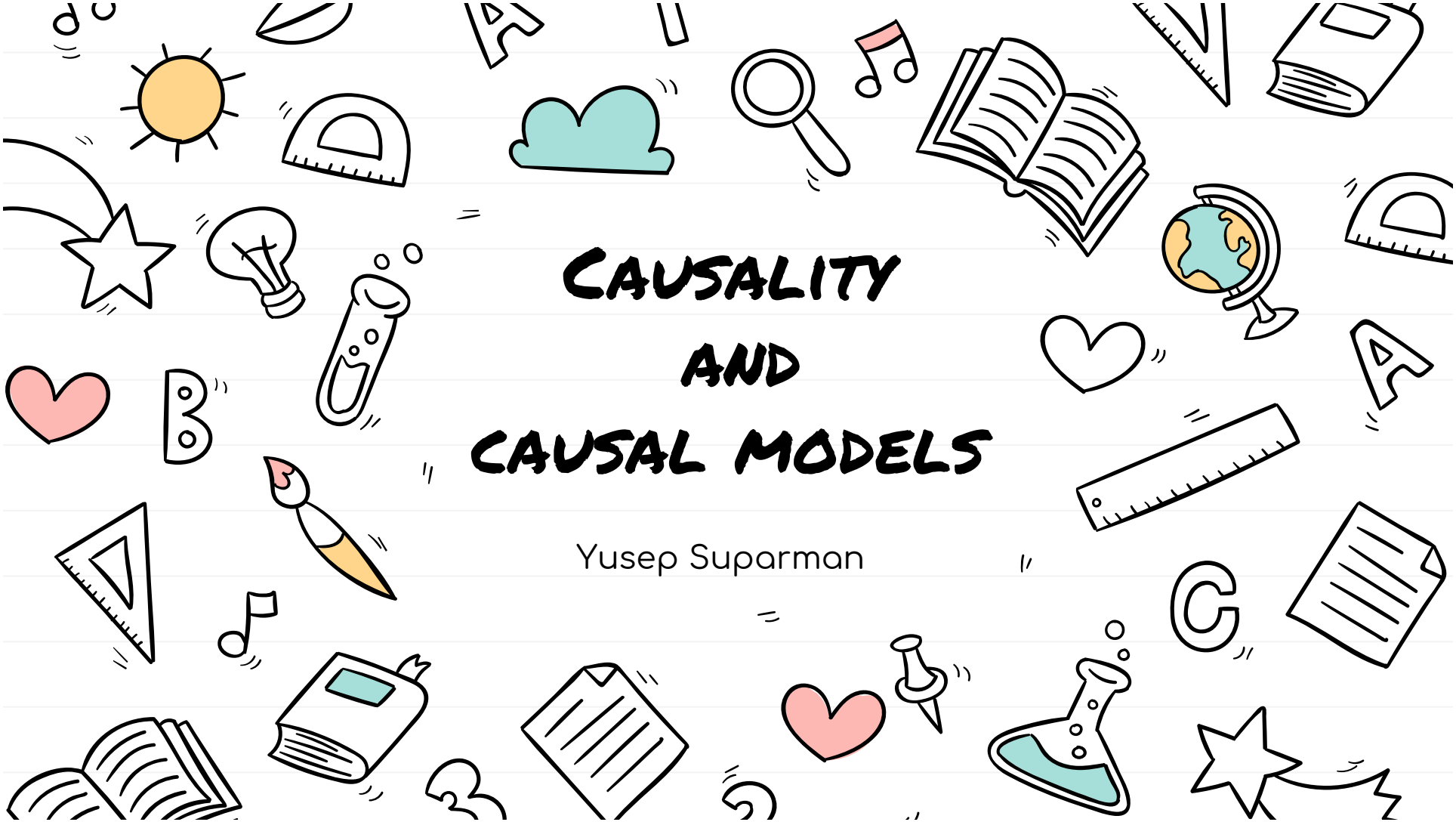


Causality and Causal Models

Yusep Suparman

Mana yang benar?

a. 9 + 5 = 2

b. 9 + 1 = 10

INTRODUCTION

- An applied research starts with a practical problem, e.g.
  - Covid-19 outbreak
  - High poverty rate
  - Bad corporate governance
  - Poor quality financial report
  - Fraudulent financial reporting
- A research is aimed at generating knowledge for solving a problem

"WHAT IS A PROBLEM?"

"PROBLEM IS A SITUATION THAT IS UNSATISFACTORY AND CAUSES DIFFICULTIES."

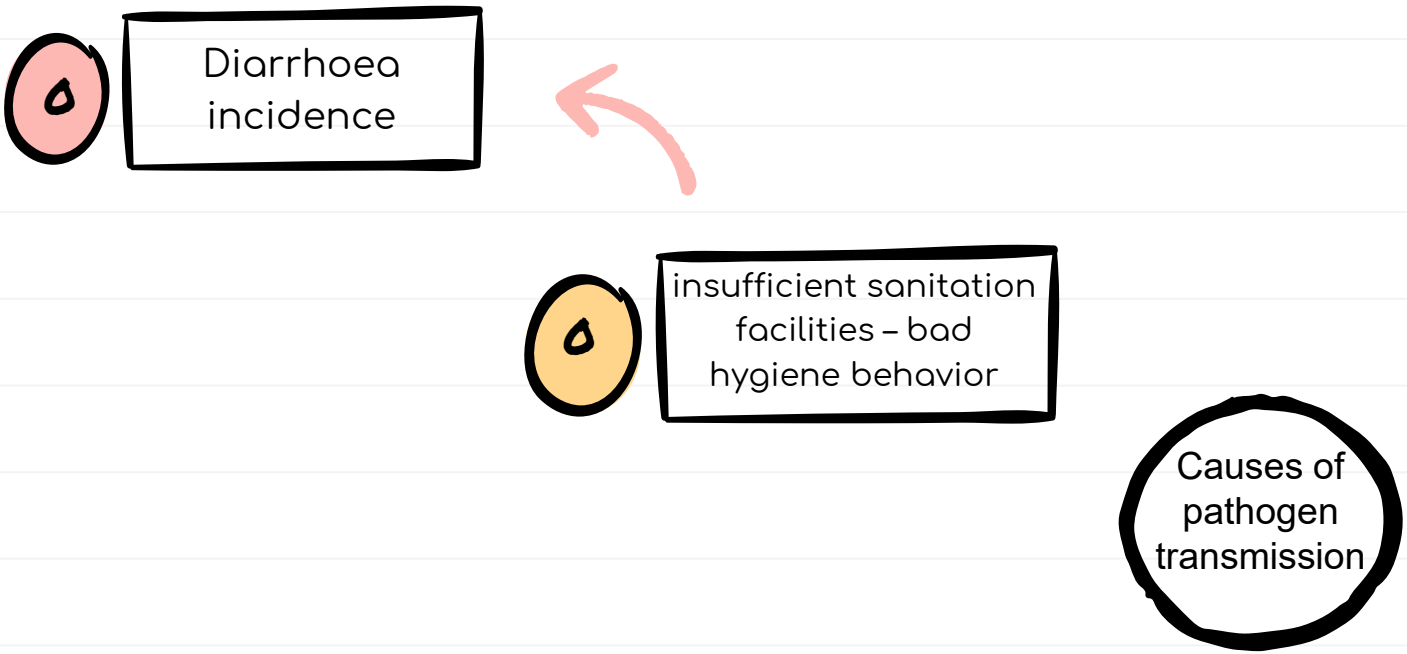
Collins Dictionary

SOLVING A PROBLEM : "CHANGING AN UNSATISFACTORY SITUATION INTO A SATIFACTORY SITUATION"

" TO CHANGE A SITUATION WE SHOULD KNOW WHAT ARE THE CAUSES OF THE SITUATION"

"BY CHANGING THE CAUSES (INTERVENTION), THE SITUATION WILL CHANGE ACCORDINGLY"

Diarrhoea incidence



Methodological aspects

- **Situations and causes are formally expressed by variables.**
- Finding the causes of a situation is identical to finding variables having causal relations to the variable representing the situation.
- Here we focus on causal relationship among variable, causality.
- In the line with data analysis, we will discuss causality's definition used in structural equation modeling.

"CONSIDER ONE VARIABLE, SAY, Y, WHICH IS ISOLATED FROM ALL INFLUENCES EXCEPT FROM A SECOND VARIABLE CALLED X. IF A CHANGE IN Y ACCOMPANIES A CHANGE IN X, THEN X IS A CAUSE OF Y"

Bollen, 1989

01

02

03

Isolation

Association

Influence direction

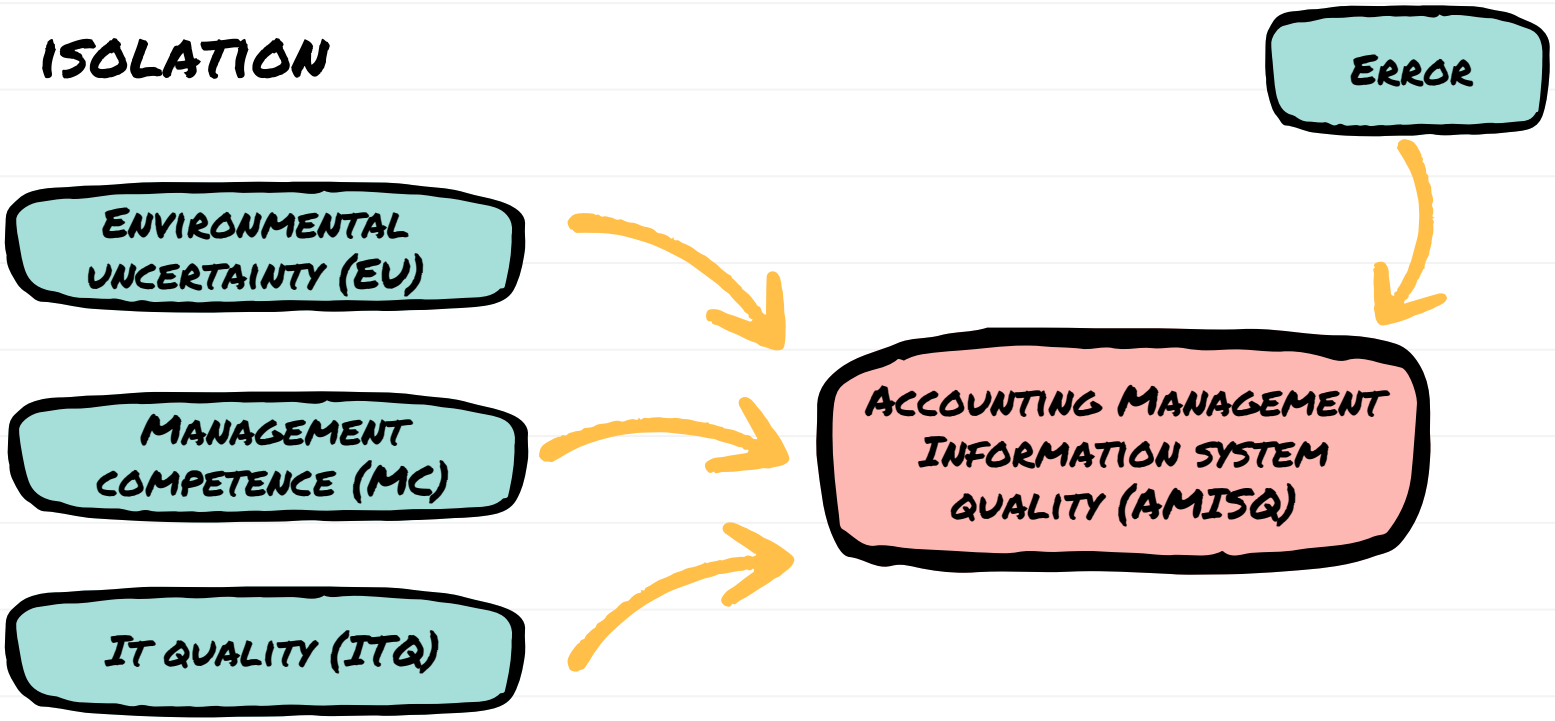
Methodological

Substantive

ISOLATION

- "y is isolated from all influences except from x"
  - Causality is defined between one variable and another one (two variables), not between one and many variables (more than two variables).
  - If there is a change in y, it comes only from the change in x, the other variables are held constant (unchanged)
- Isolation is achieved by experiment/sampling design or statistical modeling (maintaining an assumption).

ISOLATION


Error
Environmental uncertainty (EU)
Management competence (MC)
It quality (ITQ)
Accounting Management Information system quality (AMISQ)

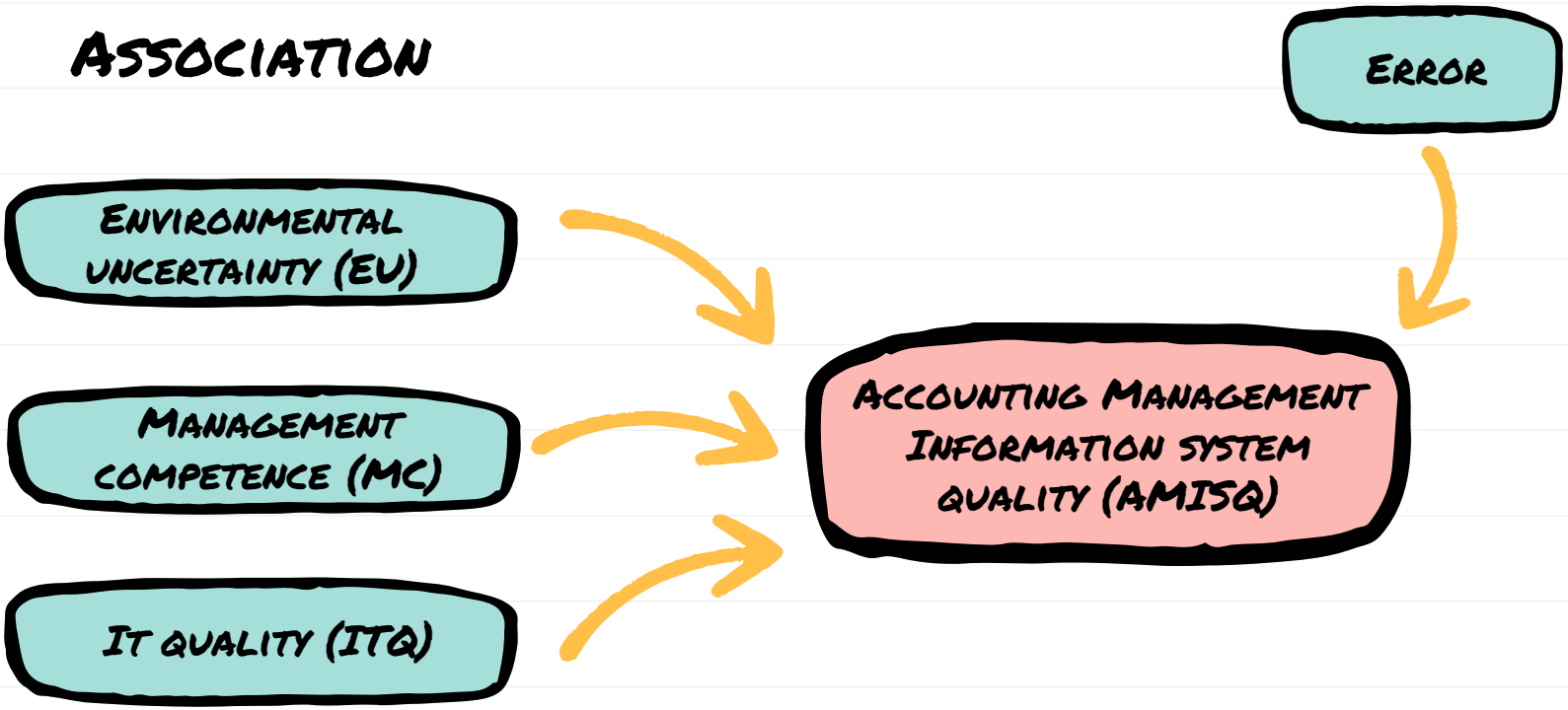
ISOLATION

- If we are interested in causal relation between MC and AMISQ
  - If we can ensure/set that all unit observations have the same level of EU and ITQ, a statistical model of MC on AMISQ is sufficient to establish causal relation between the two. Here, EU and ITQ are controlled by design. (case 1)
  - If we cannot ensure/set that all unit observations have the same level of EU and ITQ, we should put EU and ITQ into the model in addition to AMISQ. A statistic model ensures the condition of "ceteris paribus". Here, EU and ITQ are controlled by modelling. (case 2)

Assosiation

- "a change in y accompanies a change in x"
  - A statistic indicating the change in y due to the change in x, a means difference, a correlation coefficient, a regression coefficient, a path coefficient, a loading factor,
  - Non-zero change in y accompanies a change in x, indicating the presence of causal relation between x and y.

Association



Assosiation

- Case 1
  - If MC's responses are dichotomous, AMISQ-continuous, we can use two-means difference.
  - If MC's and AMISQ responses are continuous, we can use a Person's correlation or a regression coefficient
  - If both MC's and AMISQ responses are ordinal, we use a Kendall's Tau correlation or a polychoric correlation coefficient.

Assosiation

- Case 2
  - If AMISQ's responses are dichotomous, we can use a coefficient from a multiple logistic/probit/tobit regression, or multiple regression based on biserial/tethrachoric correlations.
  - If AMISQ's response are ordinal, we can use a coefficient from ordered logit/probit/tobit regression, or multiple regression based on polyserial/polychoric correlations.
  - If AMISQ responses are continues, we can use a coefficient from multiple regression.

Influence direction

- "influence from a second variable"
  - It is a substantive matter.
  - It is the domain of the expert in the research field.
  - It is nothing to do with empirical evidence or data analysis.
- Determined by literature study or critical thinking (new idea) related to the research subject.

Data analysis

- Objective:
  - Maintaining “isolation”.
  - Obtaining good association measures (statistics).
  - Conduct good inferences.
- What is a good statistics?
  - Unbias (consistent)
  - Has a minimum variance (efficient)
- What is a good inference
  - Based on a correct sampling distribution
  - Most powerful test
  - Shortest confidence interval

Data analysis

- Maintaining "isolation" by including important cause variables that are uncontrolled in experiment or sampling design into the model.
- Regression analysis is the basic statistical causal analysis. Many other statistical analysis can be expressed as a regression analysis.
- Let's focus on it to have a good understanding in statistical causal analysis.
- There are two different objectives of regression analysis: prediction – confirmation. The two have different analysis procedures.
- Causal analysis relates to the confirmation objective.

Regression analysis

- Model:

$$y_1 = \gamma_{11} x_1 + \gamma_{12} x_2 + \cdots + \gamma_{1p} x_p + \zeta_1$$

- In a regression analysis, isolation (pseudo isolation) is operationalized by assuming that the error is not correlated with all the regressor (exogeneity assumption).
- Violation of this assumption (endogeneity) results in wrong regression coefficient estimates (bias/inconsistent estimates).

Regression Analysis

- Sources of endogeneity:
  - Omitted important cause variables.
  - Measurement error in regressors
  - Simultaneity
  - Lag dependent variable as regressor (in panel or time series model)
- The sources can be first identified substantively. (conceptual framework → Model specification)
- Remedies:
  - Fix effect panel model (omitted variable)
  - Instrumental variable (endogeneity in general)
  - SEM (measurement error and simultaneity)

Regression Analysis

- The other assumptions:
  - Homoscedasticity.
  - Non-autocorrelation
  - Normality
- Violation results:
  - The first two result in estimates being non-minimum variance and inefficient, wrong inferences. Remedies: FGLS, robust standard error.
  - For a small sample size, normality violation results in wrong sampling distribution, wrong inferences. Remedies: Bootstrap, increase sample size.