ENSITE

Environmental Socio-Scientific Issues
in Initial Teacher Education

$\mathbf{M}$odule 3

COLLECTING DATA

This outline is based on the work within the project Environmental Socio-Scientific Issues in Initial Teacher Education (ENSITE). Coordination: Prof. Dr. Katja Maaß, UNIVERSITY OF EDUCATION FREIBURG, Germany. Partners: UNIVERSITEIT UTRECHT, Netherlands; ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON, Greece; UNIVERSITÄT KLAGENFURT, Austria; UNIVERZITA KARLOVA, Czech Republic; UNIVERSITA TA MALTA, Malta; HACETTEPE UNIVERSITY, Turkey; NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET NTNU, Norway; UNIVERSITY OF NICOSIA, Cyprus; INSTITUTE OF MATHEMATICS AND INFORMATICS AT THE BULGARIAN ACADEMY OF SCIENCE, Bulgaria; UNIVERZITA KONSTANTINA FILOZOFA V NITRE, Slovakia.

General overview and aim

The aim of this module is to present a basic foundation for data collection and data analysis in the light of important socio-scientific data sources to be used in lectures and seminars for mathematics and science students in the initial teacher education (ITE).

Through this module, the pre-service teachers will develop competences in planning and carrying out experiments, surveys or interviews, in collecting, preparing and analyzing data and representing it according to the rules of statistical inference. They will learn about what needs to be considered when they start planning their data collection activities with the help of experiments (e.g. using the same framework conditions, making the experiments comparable), surveys (e.g. where and how to carry out a meaningful survey, what exactly to survey, etc.), questionnaires and interviews (which items are appropriate, which are not, and how to construct items), etc. In the next step, the students will learn how to represent data. This includes lists, tables and different methods for graphical data representations, measures of the center and spread and their respective pros and cons. Last but not the least, methods for data analysis will be discussed.

Relevant topics

In this module, the emphasis is on connecting the socio-scientific issues which the ITE students research (like epidemic modelling) to data and statistics. The prospective teachers will learn how to use data and data modelling when explore and discuss specific socio-scientific issues (SSI).

Based on Module O1, the current Module O3 provides more details about data collection with respect to both learning for developing competences in dealing with environmental SSI and in teaching for acquiring the skills to educate and guide support their students in developing these competences. As part of the module, has links to Modules O2 (media data), O4 (Analyzing big data) and O5 (Decision-making), as well as to Module O8 (Beliefs on teaching environmental SSI) are provided.

Learning Outcomes

The prospective teachers are expected to acquire the following competences:

- Dealing with the essential data analysis methods learned
- Applying the data analysis methods learned in specific examples of related to environmental issues data sources
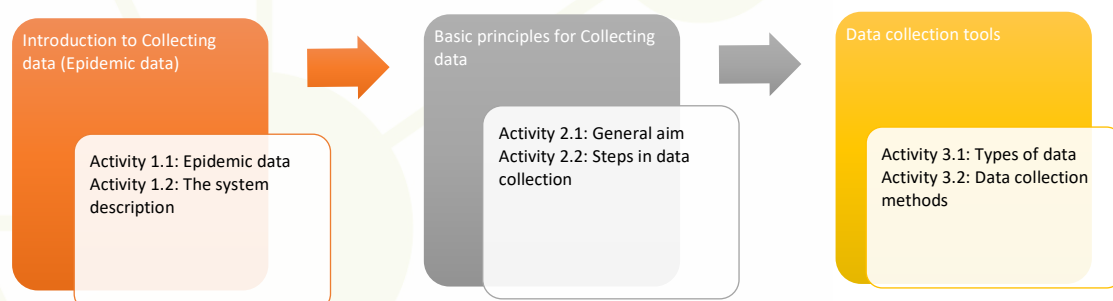
- Ability to provide meaningful reflections ~~to social life~~

- Solid knowledge on basic data analysis concepts: sample, population, descriptive statistics, statistical inference

- Skills for planning and carrying out data collection for experiments, surveys and ~~or~~ interviews

- Step-by-step dealing with data from surveys and interviews regarding their reliable collection, analysis and relevant inference

- Using different methods to represent data

- Identifying the incorrect use of survey inference in examples of political elections, drug approval, climate change, and others

- Comparing the applicability and effectiveness of various methods of dealing with environmental survey data

- Self-confidence in applying data analysis in the correct way

- Critical thinking towards media data reports followed from O2

- Adopting various strategies for decision-making

- In-depth understanding of all steps of the decision-making process from carrying it out through the evaluation of the conditions to the interpretation of results, as well as pointing out the possible consequences

Flowchart and Module plan

The module involves three sections which are structured according to the different activities they comprise. The module includes 360 minutes of sessions and 90 minutes of homework. On their side, the sessions include lecture parts, group discussions, student presentations and lab work structured in time as follows:

- Introduction to the topic through an Example about epidemic data: 90 min
- Basic principles for Collecting data: 90 min
- Application tasks (worksheets included): 180 min

| Introduction to Collecting data (Epidemic data)	Basic principles for Collecting data	Data collection tools
Activity 1.1: Epidemic data Activity 1.2: The system description	Activity 2.1: General aim Activity 2.2: Steps in data collection	Activity 3.1: Types of data Activity 3.2: Data collection methods

1. Epidemic data

1.1 General aim

Duration: 45 minutes

The future teachers share their knowledge on the COVID-19 pandemic as an unprecedented global health emergency. It offers an example of why scientific literacy, conceptualized broadly to include science concepts, the nature of scientific evidence, and ideas about how science works is so critical. We contend that it is essential for science education to support all learners in exploring complex issues like the COVID-19 pandemic. Learning experiences should help students develop science understandings associated with these issues and how science can inform solutions and personal decision-making.
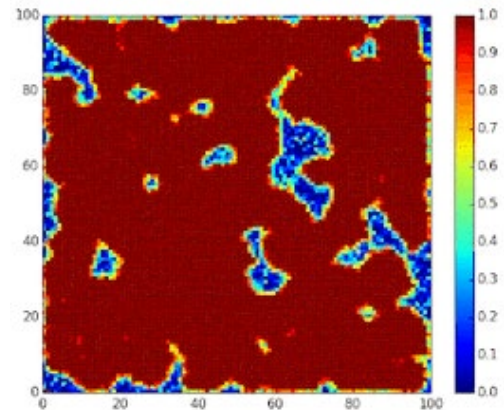

Societal issues with important connections to science such as COVID-19 are part of a broader class of problems known as Socio-Scientific Issues (SSI). Our team works to create science learning opportunities contextualized in SSI with the goal of helping students develop scientific literacy and ultimately becoming better prepared for responsibly negotiating complex challenges.

In response to COVID-19, we have partnered with several high school science teachers to create a set of SSI-based learning activities to help students learn about the science of coronavirus while engaging in epistemic practices. We use the term "practice" to denote activity that requires both disciplinary ideas and skills. Practices, such as modeling and systems thinking, are designated as "epistemic" because they are knowledge-building.

virus with a reproduction number of 2 spreads

Our goal is to create an SSI-based module using the COVID-19 pandemic as the curricular anchor. The module will help students learn about 2019-nCoV and other
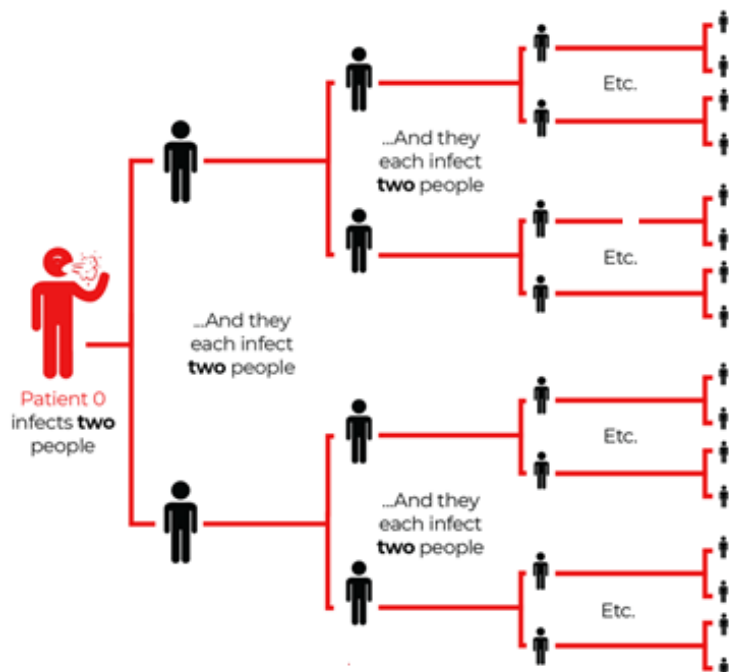
infectious viruses, how they spread, what societies can do to limit their emergence and stem their spread. A key focus of the module will be student engagement in modeling practices as they make sense of the pandemic and how to respond to it. We share the individual activities and encourage teachers to use them as they see fit (and let us know how it goes).

An SIR model is an epidemiological model that computes the theoretical number of people infected with a contagious illness in a closed population over time. The name of this class of models derives from the fact that they involve coupled equations relating the number of susceptible people S(t), number of people infected I(t), and number of people who have recovered R(t). The model studies the number of individuals in these three categories over time. Since we assume a fixed-size population, the sum of the number of individuals in all three categories is constant. So all changes in our system occur because the individuals move from one category to another.

1.2 The system description:

1) Recovered remain recovered once they enter this category. Infectives either stay infectives or they become recovered.

2) The model assumes that in each period there is a fixed part of infectives that become recovered. We name this part "healing rate" h.

3) Susceptibles either become infectives or they remain susceptibles. They become infectives by making contact with an infected individual. The new infected individuals are calculated as a part of susceptibles depending on the "contact probability" p.

Therefore, for any given susceptible individual, the probability of not becoming infected is $(1 - p)^I$. The probability of becoming infected = 1 - the probability of not becoming infected = $1 - (1 - p)^I$. Thus, the number of susceptible individuals becoming infected in any period is equal to S x $(1 - (1 - p)^I)$. The SIR diagram below shows how individuals move through each compartment in the model.

$$S_{n+1} = S_n - S_N(1 - (1 - p)^{I_n})$$
$$I_{n+1} = I_n - S_N(1 - (1 - p)^{I_n})S_n - Ih$$
$$R_{n+1} = R_n + Ih$$

Purpose: The activity presents to the SSI students a mathematical model of a spread of a virus through a human population. Like other models used by the scientists, this one is also based on specific data that describes the phenomenon of the virus spread. In this activity, students use a mathematical model, programmed within a spreadsheet (Google Sheets), to explore viral transmission and exponential growth. It enables comparisons of viruses with different reproduction numbers (R0) numerically and graphically. The activity challenges students to think through various implications of the model and to consider strategies for reducing R0 for a virus and the associated impacts and viral transmission.

2. Basic principles for Collecting data

2.1 General aim

Duration: 45 minutes

In the process of conducting a research project, the method you choose to collect your data is one of the most critical decisions you'll make. The intention is to introduce some facts about data collection in everyday life concerning SSI problems. It's important to know the different methods of data collection in order to make an objective and informed decision on the optimal collection approach to get the information you need. Also, being knowledgeable about the different collection methods enables you to better understand data that was collected in a method that is not necessarily in your interest.

Data collection is as a systematic process of gathering observations or measurements. Whether you are performing research for social or scientific purposes, data collection allows you to gain first-hand knowledge and original insights into your research problem. While methods and aims may differ between fields, the overall process of data collection remains largely the same. Before you begin collecting data, you need to consider:

- The aim of the research
- The type of data that you will collect
- The methods and procedures you will use to collect, store, and process the data.

Focus			
- Describe program model - Define purpose	**Collect data**		
- Determine use/users - Determine key questions	- Identify sources - Select methods	**Analyse & Interpret**	
- Select indicators - Determine design	- Pilot test - Set schedule	- Process data - Analyse	**Use**
	- Determine sample	- Interpret data - What did you learn? - What are the limitations?	- Share findings - Use decision making - Determine next steps

This session serves as introduction to the basic principles for data collection and data analysis in the context of environmental data sources to be used in lectures and seminars for mathematics and science students in the initial teacher education. It contributes to the achievement of the following learning outcomes:

- Raising the awareness of future teacher concerning the necessity for learning data collection principles
- Importance of the plan for carry out data collection.

ESSENTIAL QUESTIONS

What sort of methodological considerations are necessary to collect data in your educational context?

What methods of data collection will be most effective for your study?

What are the affordances and limitations associated with your data collection methods?

What does it mean to triangulate data, and why is it necessary?

As you develop an action plan for your action research project, you will be thinking about the primary task of conducting research, and probably contemplating the data you will collect. It is likely you have asked yourself questions related to the methods you will be using, how you will organize the data collection, and how each piece of data is related within the larger project.

2.2 Steps in data collection
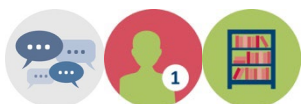
Duration: 45 minutes

In this activity, the students will realize the essential rules for collecting high-quality data that is relevant to someone's purposes.

Step 1: Define the aim of your research

Before you start the process of data collection, you need to identify exactly what you want to achieve. You can start by writing a problem statement: what is the practical or scientific issue that you want to address and why does it matter?

Next, formulate one or more research questions that precisely define what you want to find out. Depending on your research questions, you might need to collect quantitative or qualitative data.

Step 2: Choose your data collection method

Based on the data you want to collect, decide which method is best suited for your research.

- Experimental research is primarily a quantitative method.
- Interviews/focus groups and ethnography are qualitative methods.
- Surveys, observations, archival research and secondary data collection can be quantitative or qualitative methods.

Carefully consider what method you will use to gather data that helps you directly answer your research questions.

Step 3: Plan your data collection procedures

When you know which method(s) you are using, you need to plan exactly how you will implement them. What procedures will you follow to make accurate observations or measurements of the variables you are interested in? For instance, if you're conducting surveys or interviews, decide what form the questions will take; if you're conducting an experiment, make decisions about your experimental design.

Operationalization. Sometimes your variables can be measured directly while often you'll be interested in collecting data on more abstract concepts or variables that can't be directly observed. Operationalization means turning abstract conceptual ideas into measurable observations. When planning how you will collect data, you need to translate the conceptual definition of what you want to study into the operational definition of what you will actually measure.

Sampling. You may need to develop a sampling plan to obtain data systematically. This involves defining a population, the group you want to draw conclusions about, and a sample, the group you will actually collect data from. Your sampling method will determine how you recruit participants or obtain measurements for your study. To decide on a sampling method you will need to consider factors like the required sample size, accessibility of the sample, and timeframe of the data collection.

Standardizing procedures. If multiple researchers are involved, write a detailed manual to standardize data collection procedures in your study. This means laying out specific step-by-step instructions so that everyone in your research team collects data in a consistent way. This helps ensure the reliability of your data, and you can also use it to replicate the study in the future.

Creating a data management plan. Before beginning data collection, you should also decide how you will organize and store your data.

- If you are collecting data from people, you will likely need to anonymize and safeguard the data to prevent leaks of sensitive information (e.g. names or identity numbers).
- If you are collecting data via interviews or pencil-and-paper formats, you will need to perform transcriptions or data entry in systematic ways to minimize distortion.
- You can prevent loss of data by having an organization system that is routinely backed up.

Step 4: Collect the data

Finally, you can implement your chosen methods to measure or observe the variables you are interested in. To ensure that high quality data is recorded in a systematic way, here are some best practices:

- Record all relevant information as and when you obtain data. For example, note down whether or how lab equipment is recalibrated during an experimental study.
- Double-check manual data entry for errors.

- If you collect quantitative data, you can assess the reliability and validity to get an indication of your data quality.

This session serves as introduction to the basic steps in data collection. It contributes to the achievement of the following learning outcomes:

- Importance of the plan for carry out data collection in experiments, surveys or interviews

- How to deal with survey data in general: Steps needed to deal with collecting data, analysis and inference from them.

3. Data collection tools

3.1 Types of data

Duration: 45 minutes

The data collection methods have originated from a variety of disciplines, which has resulted in a variety of research frameworks to draw upon. The challenge for teachers is to develop a research plan and related activities that are focused and manageable to study. Relying on subjective knowledge enables teachers to engage more effectively as researchers in their educational context. Teachers especially rely on this subjective knowledge in educational contexts to choose data collection methodologies.

Thinking about Types of Data. Whether the research design is qualitative, quantitative or mixed-methods, it will determine the methods or ways you use to collect data. Qualitative research designs focus on collecting data that is relational, interpretive, subjective, and inductive; whereas a typical quantitative study, collects data that are deductive, statistical, and objective.

The goal of qualitative data collection is to build a complex and nuanced description of social scientific problems from multiple perspectives. The flexibility and ability to use a variety of data collection techniques encompasses a distinct stance on research. Qualitative researchers are able to capture conversations and everyday language, as well as situational attitudes and beliefs. Qualitative data collection is able to be fitted to the study, with the goal of collecting the most authentic data, not necessarily the most objective. Quantitative research depends upon structure and is bounded to find relationship among variables and units of measurement. Quantitative research helps make sense of large amounts of data. Both quantitative and qualitative research help us address education challenges by better identifying what is happening, with the goal of identifying why it is happening, and how we can address it.

Types of data. Data are organized into two broad categories: qualitative and quantitative.

Co-funded by the
Erasmus+ Programme
of the European Union

Qualitative Data: Qualitative data are mostly non-numerical and usually descriptive or nominal in nature. This means the data collected are in the form of words and sentences. Often (not always), such data captures feelings, emotions, or subjective perceptions of something. Qualitative approaches aim to address the 'how' and 'why' of a program and tend to use unstructured methods of data collection to fully explore the topic. Qualitative questions are open-ended. Qualitative methods include focus groups, group discussions and interviews. Qualitative approaches are good for further exploring the effects and unintended consequences of a program. They are, however, expensive and time consuming to implement. Additionally the findings cannot be generalized to participants outside of the program and are only indicative of the group involved.

Qualitative data collection methods play an important role in impact evaluation by providing information useful to understand the processes behind observed results and assess changes in people's perceptions of their well-being. Furthermore qualitative methods can be used to improve the quality of survey-based quantitative evaluations by helping generate evaluation hypothesis; strengthening the design of survey questionnaires and expanding or clarifying quantitative evaluation findings. These methods are characterized by the following attributes:

- they tend to be open-ended and have less structured protocols;
- they rely more heavily on interactive interviews; respondents may be interviewed several times to follow up on a particular issue, clarify concepts or check the reliability of data;
- they use triangulation to increase the credibility of their findings);
- generally their findings are not generalizable to any specific population, rather each case study produces a single piece of evidence that can be used to seek general patterns among different studies of the same issue.

Regardless of the kinds of data involved, data collection in a qualitative study takes a great deal of time. The researcher needs to record any potentially useful data thoroughly, accurately, and systematically, using field notes, sketches, audiotapes, photographs and other suitable means. The data collection methods must observe the ethical principles of research. The qualitative methods most commonly used in evaluation can be classified in three broad categories:

- In-depth interview
- Observation methods
- Document review.

Quantitative Data: Quantitative data is numerical in nature and can be mathematically computed. Quantitative data measure uses different scales, which can be classified as nominal scale, ordinal scale, interval scale and ratio scale. Often (not always), such data includes measurements of something. Quantitative approaches address the 'what' of the program. They use a systematic standardized approach and employ methods such as surveys and ask questions. Quantitative approaches have the advantage that they are cheaper to implement, are standardized so comparisons can be easily made and the size of the effect can usually be measured. Quantitative approaches however are limited in their capacity for the investigation and explanation of similarities and unexpected differences. It is important to note that for peer-based programs quantitative data collection approaches often prove to be difficult to implement for agencies as lack of necessary resources to ensure rigorous implementation of surveys and frequently

experienced low participation and loss to follow up rates are commonly experienced factors. The Quantitative data collection methods rely on random sampling and structured data collection instruments that fit diverse experiences into predetermined response categories. They produce results that are easy to summarize, compare, and generalize. If the intent is to generalize from the research participants to a larger population, the researcher will employ probability sampling to select participants. Typical quantitative data gathering strategies include:

- Experiments/clinical trials.
- Observing and recording well-defined events (e.g., counting the number of patients waiting in emergency at specified times of the day).
- Obtaining relevant data from management information systems.
- Administering surveys with closed-ended questions (e.g., face-to face and telephone interviews, questionnaires etc).
- In quantitative research (survey research), interviews are more structured than in Qualitative research. In a structured interview, the researcher asks a standard set of questions and nothing more. Face -to -face interviews have a distinct advantage of enabling the researcher to establish rapport with potential participants and therefore gain their cooperation.
- Paper-pencil-questionnaires can be sent to a large number of people and saves the researcher time and money. People are more truthful while responding to the questionnaires regarding controversial issues in particular due to the fact that their responses are anonymous.

3.2 Data collection methods
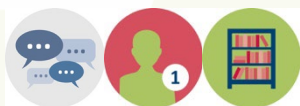
  Duration: 45 minutes

There are many ways of classifying data. A common classification is based upon who collected the data.

Primary data. Data that has been collected from first-hand-experience is known as primary data. Primary data has not been published yet and is more reliable, authentic and objective. Primary data has not been changed or altered by human beings; therefore its validity is greater than secondary data. Importance of Primary Data: In statistical surveys it is necessary to get information from primary sources and work on primary data.

Sources of Primary Data: Sources for primary data are limited and at times it becomes difficult to obtain data from primary source because of either scarcity of population or lack of cooperation. Following are some of the sources of primary data.

- **Experiments:** Experiments require an artificial or natural setting in which to perform logical study to collect data. Experiments are more suitable for chemistry, biology, medicine, and for other scientific studies. In experiments the experimenter has to keep control over the influence of any extraneous variable on the results.

- **Survey:** Survey is most commonly used method in social sciences, management, marketing and psychology to some extent. Surveys can be conducted in different methods.
- **Questionnaire:** It is the most commonly used method in survey. Questionnaires are a list of questions either open-ended or close-ended for which the respondents give answers.
- **Interview:** Interview is a face-to-face conversation with the respondent. In interview the main problem arises when the respondent deliberately hides information otherwise it is an in depth source of information.
- **Observations:** Observation can be done while letting the observing person know that s/he is being observed or without letting him know. Observations can also be made in natural settings as well as in artificially created environment.

Advantages of Using Primary Data:

- The investigator collects data specific to the problem under study.
- There is no doubt about the quality of the data collected (for the investigator).
- If required, it may be possible to obtain additional data during the study period.

Secondary data. Data collected from a source that has already been published in any form is called as secondary data. The review of literature in any research is based on secondary data. It is collected by someone else for some other purpose (but being utilized by the investigator for another purpose). For examples, Census data being used to analyze the impact of education on career choice and earning. Common sources of secondary data for include censuses, organizational records and data collected through qualitative methodologies or qualitative research. Secondary data is essential, since it is impossible to conduct a new survey that can adequately capture past change and/or developments.

Method	When to use	How to collect data
Experiment	To test a causal relationship	Manipulate variables and measure their effects on others
Survey	To understand the general characteristics or opinions of a group of people	Distribute a list of questions to a sample online, in person or over-the-phone.
Interview/focus group	To gain an in-depth understanding of perceptions or opinions on a topic	Verbally ask participants open-ended questions in individual interviews or focus group discussions.
Observation	To understand something in its natural setting.	Measure or survey a sample without trying to affect them.

Secondary data collection	To analyze data from populations that you can't access first-hand	Find existing datasets that have already been collected, from sources such as government agencies or research organizations

Based on the data you want to collect, decide which method is best suited for your research.

Carefully consider what method you will use to gather data that helps you directly answer your research questions.
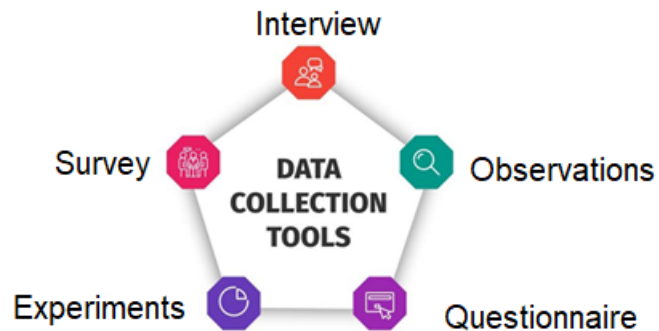
Materials and resources

- J. Collins and N. Abdelal Spread of Disease. https://calculate.org.au/wp-content/uploads/sites/15/2018/10/spread-of-disease.pdf Spread of Disease Video
- Steve Strain (Department of Biology, Slippery Rock University of Pennsylvania) "SIR Epidemic Dynamics" http://demonstrations.wolfram.com/SIREpidemicDynamics/ Wolfram Demonstrations Project

Access to computers for internet research and collaborative work

References

References
Sapsford, R., & Jupp, V. (Eds.). (1996). *Data collection and analysis*. Sage.
Peter Stopher (2012). - Collecting, Managing, and Assessing Data Using Sample Surveys-Cambridge University Press
Maasz, J., & O'Donoghue, J. (Eds.). (2011). *Real-world problems for secondary school mathematics students*. Springer Science & Business Media..

Ledder, G., Homp, M. (2020). <u>Materials for Teaching the SIR Epidemic Model</u>. QUBES Educational Resources. <u>doi:10.25334/4H67-PT17</u>

Model for Viral Spread

https://docs.google.com/spreadsheets/d/1f7wkHPMYQfGYaeiOQxElx0OIoaFytKpwLyNGhm5tqkg/edit#gid=0

Further readings

- Alvarez, M.M., González-González, E. & Trujillo-de Santiago, G. Modeling COVID-19 epidemics in an Excel spreadsheet to enable first-hand accurate predictions of the pandemic evolution in urban areas. *Sci Rep* **11,** 4327 (2021). https://doi.org/10.1038/s41598-021-83697-w

Assessment

Question 1

1a) Suppose the reproduction number (R0) for COVID-19 may be as low as 1.5 or as high as 3.5. How many people would we expect to be infected after 5 rounds of transmission if the low R0 is accurate?

Answer: *If you start with 1 infected person, after 5 rounds of transmission about 8 people will be infected.*

1b) How many people would we expect to be infected after 5 rounds of transmission if the high R0 is accurate?

Answer: *If you start with 1 infected person, after 5 rounds of transmission about 525 people will be infected.*

1c) Why does the accuracy of the R0 value matter?

Answer: *Answers may vary. One possible answer: Because each round of transmission is exponential, a small change in R0 results in a large difference in the number of infected people.*

1d) Why do you think there is such a big range for the R0 value for COVID-19 at this point in time?

Answer: *Answers may vary. One possible answer: We are still learning more about the spread of Covid-19. There are a number of additional factors that could impact the rate of transmission such as social distancing and wearing masks.*

Question 2

2a) R0 for the common flu is estimated at 1.3. If we assume R0 for COVID-19 is 3, how many more people can we expect to be infected by COVID-19 compared to the flu after 10 rounds of transmission?

Answer: *Assuming you start with 1 infected, the common flu would infect 14 people total. Covid-19 would infect 59,049. About 59,035 more would be infected by Covid-19.*

2b) How many rounds of transmission of the flu would it take for flu to infect the number of people infected by COVID-19 after 10 rounds of transmission?

Answer: *About 42 rounds of transmission.*

2c) If you had to calculate the number infected after 3 rounds of transmission by hand

(without the use of the model), explain how you would make the calculation.
Answer: *First round: You multiply the initial number of infected by the R0. Second round: You multiply the total infected from the first round by the R0. Third round: You multiple the total infected from the second round by R0.*

Question 3
3a) This model helps us to think about the spread of disease and can help us to compare the rate of spread of different viruses. What other aspects of different viruses should we consider when comparing the "intensity" of viral outbreaks?
Answer: *Answer may vary. Possible answers: virus mutations, replication rate of viruses in the body, how viruses enter the body.*
3b) What are limitations of this model?
Answer: *Answers may vary. Possible answers: the model does not factor in social distancing or other efforts to slow the spread, the model is an estimation, the model cannot represent any changes in R0 overtime.*

Question 4
4a) What happens to a viral outbreak when R0 for the virus fall below 1?
Answer: *The rate of transmission will be decrease instead of increase.*
4b) What things could be done to lower the virus R0?
Answer: *Answers may very. Possible answers: hand washing, wearing masks, social distancing, quarantining.*
4c) If we assume that 100,000 individuals are infected by virus but we were able to drive its R0 down to 0.2, how long (in rounds of transmission) would it take to halt the spread of the virus?
Answer: *8 rounds.*