Modeling and Simulation Project: Evacuation

Professor: Alexis Drogoul

Vu Trung Dung — dungvt2440071@usth.edu.vn

January 11, 2025

Abstract

Populations are increasingly vulnerable to disastrous natural or technological events, as demographic and urban growth lead to greater exposures of goods and people. Hanoi, for example, is particularly hard hit by flooding. Some districts on the banks of the Red River are also threatened by potential dike breaching. In the event of a levee failure, it is important to be able to evacuate the population living in these areas before the water arrives.

1 Introduction

1.1 Problem Statement

The goal of this project is to build an agent-based model of people evacuation, to have a better understanding of the evacuation process and to be able to test different evacuation strategies. The question to be answered is: **How to better manage the pedestrian evacuation of a population on a beach in a tsunami context?**

The goal is to optimize the evacuation process according to the following criteria:

- Minimize the time people spend wandering on the roads to inform others about the threat.

- Maximize the number of people informed and evacuated.

- Minimize the total evacuation time as well.

1.2 Project Extensions

In this model, flooding will not be modeled by itself, just the behavior of residents in the face of the threat. People will only evacuate if they have been informed of the imminent risk of flooding. At the start of the simulation, we assume that all residents are located in their own homes and only 10% of the population (randomly chosen) will be aware of this information. Once informed, people will evacuate to the shelter (the largest building in the area). A person observing someone evacuating (at a distance of less than 10m) will have a probability of 0.1 of evacuating in turn.

1. **Extension 1**: Only 10% of the population go directly to the shelter when informed of the risk. The other will move to random buildings to inform other residents and search for the shelter.

2. **Extensions 2**: Implement the multiple modalities of evacuation (car, motocycle, walking) and the possibility of blocking roads.

3. **Extensions 3**: Experiment with different evacuation strategies of informing the threat to the 10% of the population: those furthest from shelter, those closest to the shelter, or randomly selected.

2   Implementation

In this section, I will describe the implementation of the agent-based model for the evacuation process.

2.1   Base Model

The base model include the following **species**:

- **Inhabitant**: Represented as people that can move on the map, spread threat information, and evacuate to the shelter.

- **Buildings**: Represented as static objects where people can stay or move to for shelter.

- **Roads**: Represented as paths that people can use to move around the area.

- **Shelter**: Represented as the largest building in the area where people can evacuate to.

The people is randomly placed in their homes at initialization. To make it more realistic, I suppose that there are at maximum 5 people in each house. When the threat is announced, I randomly choose 10% of the population to be informed. People who are informed have 10% change to evacuate directly to the shelter, other will wander around to inform others and search for the shelter. In GAMA implementation, **inhabitant** species have included the skills **moving**, which allow them to move around through the roads. They realize the shelter if they found it at 20m distance. They stop moving when they reach the shelter and completely evacuate.

In my implementation, I suppose that **people don't revisit the same place they have been before**, so they will not wander around the same place, to increase the chance of finding the shelter. This is done by maintaining a list of visited places for each inhabitant.

The tsunami is not modeled in this base model, there is only the alert time before the tsunami arrives and the evacuation process. When tsunami strikes, it's often too late to inform and evacuate people. So the another way to express the question: **How effective are different evacuation processes in maximizing the number of evacuees**

within a limited time frame?

According to that, the simulation start with the alert time, and the evacuation process will be stopped after a certain time or when all people have evacuated.

Note that to save the resources, I will destroy the inhabitant agents when they reach the shelter, after counting them as evacuated.

The most important part of the inhabitant behaviour is the **find_shelter** reflex function, which is responsible for the evacuation process. The function is described as follows:

```
GAMA

reflex find_shelter when: target = nil and is_evacuating {
    if (shelter distance_to self < 1#m) {
        number_evacuted_people <- number_evacuted_people + 1;
        total_time_in_roads <- total_time_in_roads + (current_date -
            start_evacuating_date);
        total_evacuation_time <- total_evacuation_time + time;
        do die;
        return;
    }

    building target_building <- known_shelter ? shelter : nil;
    if (target_building = nil) {
        write("randomly find another shelter");
        if (shelter distance_to self < 20#m) {
            target_building <- shelter;
        }
        target_building <- one_of((building - visited_buildings) where each.is_safe
            );
    }
    visited_buildings << target_building;
    target <- any_location_in(target_building);
}
```

In this code, the **is_evacuating** and **known_shelter** are the agent's attributes that indicate the evacuation status and the knowledge of the shelter location. The **visited_buildings** is the list of buildings that the agent has visited, to avoid revisiting the same place. The **number_evacuted_people** is the global variable that counts the number of people evacuated. The **total_time_in_roads** and **total_evacuation_time** are the global variables that count the total time people spend wandering on the roads and the total evacuation time. Finally, the built-in attribute **target** to set to any location in the **target_building** to move to that location.

2.2   Extension 1

As describe in the previous code, each inhabitant include the **known_shelter** attribute to indicate if they know the shelter location. This is randomly set to 10% of the population at the beginning of the simulation.

2.3 Extension 2

In this extension, I simulate the road heavily congested by cars and motorcycles, which block the way for pedestrians. Each inhabitant has their own mobilities which set randomly at the beginning of the simulation. The mobilities include **car**, **motorcycle**, and **walking**.

Here is the code represented the **road_weights** of the roads in the simulation:

```
GAMA

global {
    reflex update_speed {
        road_weights <- road as_map (each::each.shape.perimeter / each.speed_rate);
    }
}

species road {
    float capacity <- 1 + shape.perimeter/10;
    float total_traffic_weight <- 0.0
        update: sum((inhabitant at_distance 1) collect each.traffic_weight);
    float speed_rate <- 1.0 update:  exp(-total_traffic_weight/capacity) min: 0.1;
}
```

2.4 Extension 3

In this extension, I experiment with different evacuation strategies of informing the threat to the 10% of the population:

- **Furthest**: Inform the threat to the 10% of the population who are furthest from the shelter.

- **Closest**: Inform the threat to the 10% of the population who are closest to the shelter.

- **Random**: Inform the threat to the 10% of the population who are randomly selected.

In GAMA, I simply define the different strategies by setting the **known_shelter** attribute of the different selected inhabitant agents. The code is as follows:

```GAMA                                                      5
if (initial_inform_strategy = "random") {
    write("Strategy: Random");
    ask nb_informing_people among inhabitant {
        do evacuate();
    }
} else if (initial_inform_strategy = "furthest") {
    write("Strategy: Furthest");
    ask nb_informing_people first (inhabitant sort_by -distance_to(shelter, each))
        {
        do evacuate();
    }
} else {
    write("Strategy: Closest");
    ask nb_informing_people first (inhabitant sort_by distance_to(shelter, each)) {
        do evacuate();
    }
}
```

The experiment of different strategies I conducted will be presented in the next section.

3   Experiments

3.1   Pre-Definitions

To have better presentation of the simulation, I define different colors for the different evacuate status of the inhabitant agents, which are:

- **Evacuating/Informed**: Blue

- **Uninformed**: Yellow

And the different shape for the different mobilities of the inhabitant agents, which are:

- **Car**: Square.

- **Motorcycle**: Triangle.

- **Walking**: Circle.

All of theses are showed in the Fig 1.

3.2   Parameters

The parameters of the simulation are set as follows (Fig 2):

- **Inform Strategy**: one of {random, furthest, closest}

- **Number of People**: default 1000

- **Alert Time**: the time before the tsunami arrives, default 120 minutes.
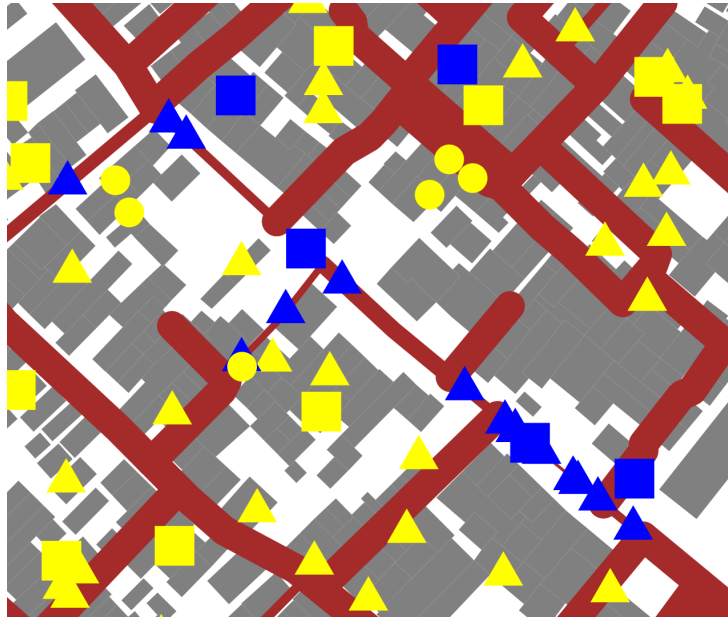
Figure 1: Different mobilities of the inhabitant agents: Car (Square), Motorcycle (Triangle), Walking (Circle). Colors represent the evacuate status: Blue (Evacuating/Informed), Yellow (Uninformed).
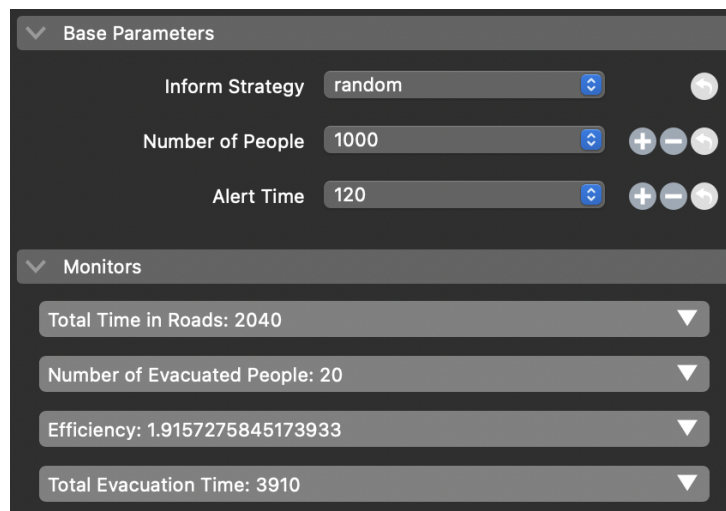

Figure 2: Parameters and Metrics of the simulation.

3.3 Metrics

To monitor and evaluate the evacuation process, I define the following metrics:

- **Number of Evacuated People**: The number of people who have evacuated to the shelter before the tsunami arrives.

- **Total Time in Roads**: The total time people spend wandering on the roads to inform others and search for the shelter. To be specified, the time is counted from the time the inhabitant is informed and start evacuating until they reach the shelter.

- **Total Evacuation Time**: The time from the alert time until the last inhabitant evacuates to the shelter.

- **Efficiency**: The efficiency of the strategy.

The **Total Time in Roads** can be calculated by the formula:

$$\text{Total Time in Roads} = \sum_{i=1}^{N} \text{T\_roads}_i$$

$$\text{T\_roads}_i = \begin{cases} \text{t\_evacuated}_i - \text{t\_informed}_i & \text{if the inhabitant reaches the shelter} \\ \text{t\_current} - \text{t\_informed}_i & \text{if the inhabitant is still on the road} \end{cases}$$

where t_evacuated_i is the time the inhabitant i reaches the shelter and t_informed_i is the time the inhabitant i is informed and start evacuating, and N is the total of informed people.

The topic suggests the **Efficiency** of the strategy, which can be calculated by the formula:

$$\text{Efficiency} = \frac{\text{Total Evacuation Time}}{\text{Total Time in Roads}}$$

But after conducting the simulation, I found that this is not a good metric to evaluate the strategy. In the worst case, when the simulation stops because of running out of time, the total evacuation time equals the total simulation runtime, the total time spent on the roads could still be low if only 10% of people were informed initially and went directly to the shelter without informing others. This scenario would yield a highest efficiency number despite being the worst strategy.

To correct this, I suggest the **Efficiency** metric to be calculated by the formula:

$$\text{Efficiency} = \frac{\text{Total Evacuated people} \times \text{Alert Time}}{\text{Total Time in Roads} + 1}$$

This metric will give a better evaluation of the strategy, as it tries to minimize the time people spend wandering on the roads to inform others and search for the shelter, and maximize the number of people informed, resulting in the total number of people evacuated. The +1 is added to avoid the division by zero. The scaling of the Alert Time is

added to make the division to be in the same unit.

Note that the chance of people successfully finding the shelter is random, and can't be controlled by the strategy. **The strategy only affects the number of people informed and the time they spend wandering on the roads, which possibility increases the chance of successfully evacuated people**.

3.4 Results

In this experiments, I conducted the simulation with different strategies and alert times. I also made the comparison between the different strategies by running them in parallel. About the map, I used the map of **Hanoi and the Red River** in the previous exercises. The shelter in this map is the one in the right-bottom corner (colored in green in the Fig 3).

However, this building is too far from the others, which makes the evacuation process more difficult, since it's hard to people to realize it at distance 20m while wandering around. I made a change by selecting the building in the center-bottom of the map as the shelter. It is presented in red color in the Fig 3. A quick experiment showed that in the settings of 3000 people and 2 hours of alert time, the number of evacuated people is increased to 298 compared to 248 of the previous shelter. Since the new shelter is more accessible, the results of comparison between the strategies are more reliable, as they eliminate the luck factor involved in finding the shelter and focus on the efficiency of the strategies.

A single simulation of the evacuation process is presented in the Fig 4.

The comparison of the different strategies in number of evacuated people is presented in the Fig 5. The experiment is conducted with 2000 and 4000 people, 2 hours alert time among the different strategies. The results show that, at the settings of large number of people, the **closest** strategy is the one that save the most people, followed by the **furthest** strategy, and the **random** strategy is the least efficient. At smaller number of people, the three strategies have similar results. In all cases, the **random** strategy introduces the fastest evacuation process at the beginning. These findings can be explained by some reasons:

- The **random** strategy is the fastest at the beginning because the people are randomly selected, which means they are distributed evenly across the map, so the people can start finding the shelter from diverse locations, which reduces the traffic congestion and the time people spend wandering on the roads, but the chance of people finding the shelter isn't increased.

- The **closest** strategy introduces the largest number of people saved because the people closest to the shelter have chance to inform others closed to the shelter as well, which increases the chance of people finding the shelter.

Next, I conducted the batch exploration to examine which strategy is the most efficient, which determined in the previous section. The results are presented in the Fig **??**.

Figure 3: The GIS data of Hanoi and the Red River. The largest building is in the bottom-right corner (colored in green). The shelter is in the center of the map (colored in red).
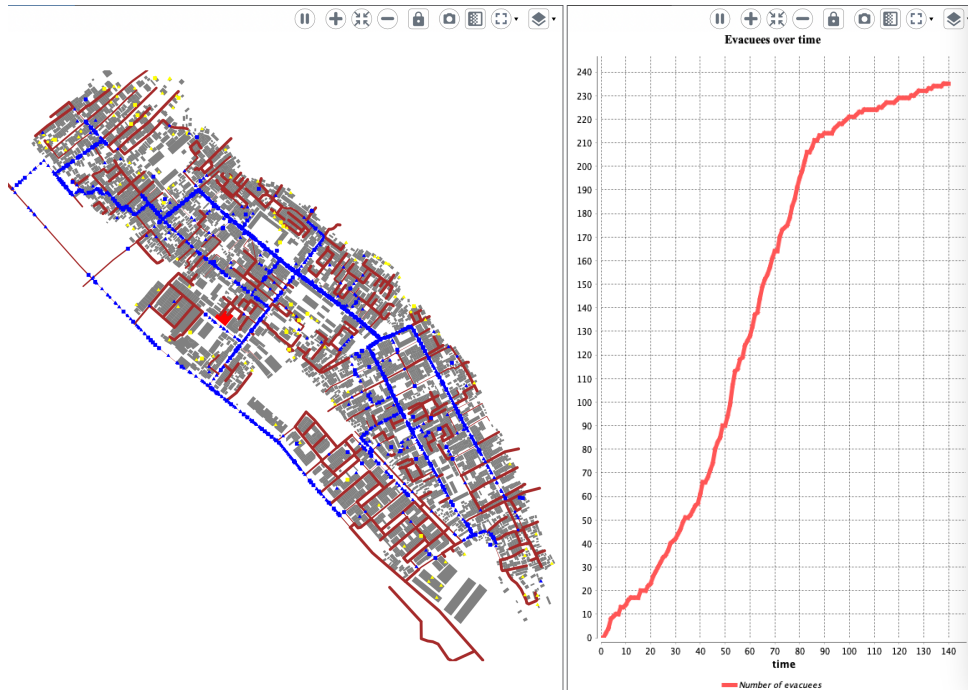

Figure 4: The evacuation simulation with 1000 people, **random** strategy and 2 hours evacuate time. The right chart is the total number of evacuated people over time. The red building in the right-bottom corner is the shelter.
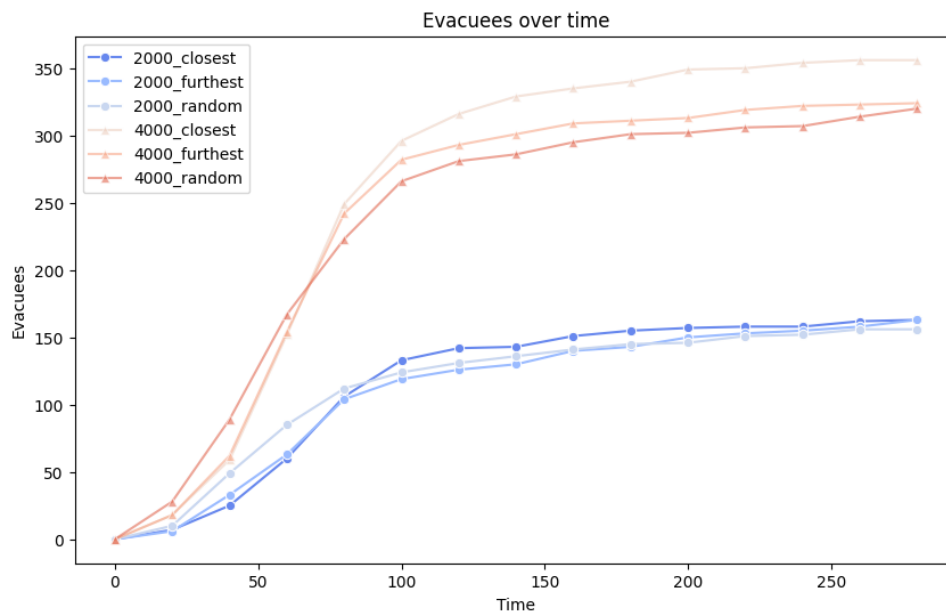
Figure 5: The comparison of the different strategies in number of evacuated people over time.

4 Discussion

Appendix A

5 General Information

5.1 What is the title of the project?

Example
Evaluating methods for the analysis of pre–post measurement experiments

Answer:

5.2 Who are the current and future project contributors?

Example
Björn S. Siepe, František Bartoš, and Samuel Pawel

Answer:

5.3 Provide a description of the project.

Explanation: This can also include empirical examples that will be analyzed within the same project, especially if the analysis depends on the results of the simulation.

> **Example**
>
> We will investigate the performance of different methods for analyzing data from pre–post measurement experiments. We will conduct a single simulation study varying the treatment effect and the pre–post measure correlation. We will compare three different methods (ANCOVA, change score analysis, and post score analysis) using power and type I error rate related to the hypothesis test of no effect, and bias related to the effect estimate (in an actual simulation study aimed at evaluating estimation of the effect size, a performance measure assessing variance, i.e., empirical standard error, would be also recommended).

Answer:

5.4 Did any of the contributors already conduct related simulation studies on this specific question?

Explanation: This includes preliminary simulations in the context of the current project.

> **Example**
>
> We did not conduct previous simulation studies for pre–post measurement experiments but we were inspired by the previous literature on the topic (**vickers2001use**; **senn2006change**; **van2013ancova**; **clifton2019correlation**; **ludtke2023ancova**).

Answer:

6 Aims

6.1 What is the aim of the simulation study?

Explanation: The aim of a simulation study refers to the goal of the research and shapes subsequent choices. Aims are typically related to evaluating the properties of a method (or multiple methods) with respect to a particular statistical task. Possible tasks include 'estimation', 'hypothesis testing', 'model selection', 'prediction', or 'design'. If possible, try to be specific and not merely state that the aim is to 'investigate the performance of method X under different circumstances'.

> **Example**
>
> The aim of the simulation study is to evaluate different methods for analyzing data from pre–post measurement experiments with respect to their hypothesis testing and estimation characteristics.

Answer:

7 Data-Generating Mechanism

7.1 How will the parameters for the data-generating mechanism (DGM) be specified?

Explanation: Answers include 'parametric based on real data', 'parametric', or 'resampled'. Parametric based on real data usually refers to fitting a model to real data and using the parameters of that model to simulate new data. Parametric refers to generating data from a known model or distribution, which may be specified based on theoretical or statistical knowledge, intuition, or to test extreme values. Resampled refers to resampling data from a certain data set, in which case the true data-generating mechanism is unknown. The answer to this question may include an explanation of from which distributions (with which parameters) values are drawn, or code used to generate parameter values. If the DGM parameters are based on real data, please provide information on the data set they are based on and the model used to obtain the parameters. Also, indicate if any of the authors are already familiar with the data set, e.g., analyzed (a subset of) it.

Example

In each simulation repetition, we generate $n = 50$ pre–post measurements in the control group ($g = $ control) and $n = 50$ pre–post measurements in the experimental group ($g = $ exp) from a bivariate normal distribution

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ \mu_{g,2} \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \tag{1}$$

where the first argument of the normal distribution is the mean vector and the second argument the covariance matrix. The numerical subscript 1 indicates measurement time 'pre' and 2 indicates 'post'. The parameter $\mu_{g,2}$ denotes the post-treatment mean. It is fixed to zero in the control group ($\mu_{\mathrm{control},2} = 0$), whereas it is varied across simulation conditions in the experimental group. The parameter ρ denotes the pre–post correlation and is also varied across simulation conditions.

Answer:

7.2 What will be the different factors of the data-generating mechanism?

Explanation: A factor can be a parameter/setting/process/etc. that determines the data-generating mechanism and is varied across simulation conditions.

Example

We will vary the following factors:

- the post-treatment mean in the experimental condition $\mu_{\mathrm{exp},2}$

- the pre–post measurement correlation ρ

Answer:

7.3   If possible, provide specific factor values for the DGM as well as additional simulation settings.

Explanation: Also include information about settings that are held constant across simulation conditions. This may include a justification of the chosen values and settings.

Example

We will use the following values for our data-generating mechanism:

- $\mu_{\exp,2} \in \{0, 0.2, 0.5\}$

- $\rho \in \{0, 0.5, 0.7\}$

We selected these specific values for the post-treatment mean in the experimental condition as they correspond to the conventions for no, small, and medium standardized mean difference effect sizes in psychology (**Cohen1988**) and pre–post measurement correlations that correspond to no, one quarter, and approximately one half of the shared variance. Based on our experience, these parameter values are relevant for empirical research while covering a sufficiently large range to allow us to observe possible differences between the examined methods. For simplicity of the example, we consider only a single sample size, namely, $n = 50$ per group.

Answer:

7.4   If there is more than one factor: How will the factor levels be combined and how many simulation conditions will this create?

Explanation: Answers include 'fully factorial', 'partially factorial', 'one-at-a-time', or 'scattershot'. Fully factorial designs are designs in which all possible factor combinations are considered. Partially factorial designs denote designs in which only a subset of all possible factor combinations are used. One-at-a-time designs are designs where each factor is varied while the others are kept fixed at a certain value. Scattershot designs include distinct scenarios, for example, based on parameter values from real-world data.

Example

We will vary the conditions in a fully factorial manner. This will result in
3 (post-treatment mean in experimental group) × 3 (pre–post measurement correlation) = 9 simulation conditions.

Answer:

8 Estimands and Targets

8.1 What will be the estimands and/or targets of the simulation study?

Explanation: Estimands and other targets jointly refer to the practical aims of the compared methods. For example, an estimand/target could be the true effect size estimated by the compared methods (**Siepe2024**). Please also specify if some targets are considered more important than others, i.e., if the simulation study will have primary and secondary outcomes.

Example

Our primary target is the null hypothesis of no difference between the outcomes of the control and treatment groups. Our secondary estimand is the treatment effect size defined as the expected difference between the control and the experimental group measurements at time-point two

$$\mathrm{E}(Y_2 \mid g = \exp) - \mathrm{E}(Y_2 \mid g = \mathrm{control}),$$

for which the true value is given by the parameter $\mu_{\exp,2}$ for the considered data-generating mechanisms.

Answer:

9 Methods

9.1 How many and which methods will be included and which quantities will be extracted?

Explanation: Be as specific as possible regarding the methods that will be compared, and provide a justification for both the choice of methods and their model parameters. This can also include code which will be used to estimate the different methods or models in the simulation with all relevant model parameters. Setting different prior hyperparameters might also be regarded as using different methods. Where package defaults are used, state this. Where they are not used, state what values are used instead.

Example

We will compare the following methods:

1) **ANCOVA** (ANalysis of COVAriance): A regression of the post-treatment measurement using the pre-treatment measurement and the treatment indicator as covariates, which is specified in R as

```
lm(post ~ pre + treatment)
```

2) **Change score analysis**: A regression of the difference between post-treatment and pre-treatment measurement using the treatment indicator as covariate, which is specified in R as

```
lm(post ~ offset(pre) + treatment)
```

3) **Post score analysis**: A regression of the post-treatment measurement using the treatment indicator as covariate, which is specified in R as

```
lm(post ~ treatment)
```

Both change score and post score ANOVA can be seen as a special case of ANCOVA. Change score analysis fixes the `pre` coefficient to 1 (using the `offset()` function) and post score analysis omits the `pre` variable from the model (effectively fixing its coefficient to 0).

From each fitted model, we will extract the estimated treatment effect, the associated standard error, and the associated two-sided t-test p-value for the null hypothesis of no effect. A rejection of the null hypothesis will be defined by a p-value less than the conventional threshold of 0.05.

Answer:

10 Performance Measures

10.1 Which performance measures will be used?

Explanation: Please provide details on why they were chosen and on how these measures will be calculated. Ideally, provide formulas for the performance measures to avoid ambiguity. Some models in psychology, such as item response theory or time series models, often contain multiple parameters of interest, and their number may vary across conditions. With a large number of estimated parameters, their performance measures are often combined. If multiple estimates are aggregated, specify how this aggregation will be performed. For example, if there are multiple parameters in a particular condition, the mean of the individual biases of these parameters or the bias of each individual parameter

may be reported.

Our primary performance measures are the type I error rate (in conditions where the true effect is zero) and the power (in conditions where the true effect is non-zero) to reject the null hypothesis of no difference between the control and treatment condition. The null hypothesis is rejected if the p-value for the null hypothesis of no effect is less than or equal to the conventional threshold of 0.05. The rejection rate (the type I error rate or the power, depending on the data generating mechanism) is estimated by

$$\widehat{\text{RRate}} = \frac{\sum_{i=1}^{n_{\text{sim}}} 1(p_i \leq 0.05)}{n_{\text{sim}}}$$

where $1(p_i \leq 0.05)$ is the indicator of whether the p-value in simulation i is equal to or less than 0.05. We use the following formula to compute the MCSE of the rejection rate

$$\text{MCSE}_{\widehat{\text{RRate}}} = \sqrt{\frac{\widehat{\text{RRate}}(1 - \widehat{\text{RRate}})}{n_{\text{sim}}}}.$$

Our secondary performance measure is the bias of the treatment effect estimate. It is estimated by

$$\widehat{\text{Bias}} = \frac{\sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i}{n_{\text{sim}}} - \theta$$

where θ is the true treatment effect and $\hat{\theta}_i$ is the effect estimate from simulation i. We compute the MCSE of the estimated bias with

$$\text{MCSE}_{\widehat{\text{Bias}}} = \frac{S_{\hat{\theta}}}{\sqrt{n_{\text{sim}}}}$$

where $S_{\hat{\theta}} = \sqrt{\sum_{i=1}^{n_{\text{sim}}} \{\hat{\theta}_i - (\sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i / n_{\text{sim}})\}^2 / (n_{\text{sim}} - 1)}$ is the sample standard deviation of the effect estimates.

Answer:

10.2 How will Monte Carlo uncertainty of the estimated performance measures be calculated and reported?

Explanation: Ideally, Monte Carlo uncertainty can be reported in the form of Monte Carlo Standard Errors (MCSEs). Please see **Siepe2024<empty citation>** and **Morris2019<empty citat** for a list of formulae to calculate the MCSE related to common performance measures, more accurate jackknife-based MCSEs are available through the `rsimsum` (**Gasparini2018**) and `simhelpers` (**Simhelpers2022**) R packages, the `SimDesign` (**Chalmers2020**) R package can compute confidence intervals for performance measures via bootstrapping.

Monte Carlo uncertainty can additionally be visualized using plots appropriate for illustrating variability, such as MCSE error bars, histograms, boxplots, or violin plots of performance measure estimates, if possible (e.g., bias).

Example
We will report Monte Carlo uncertainty in tables (MCSEs next to the estimated performance measures) and in plots (error bars with ± 1MCSE around estimated performance measures). We will use the formulas provided in **Siepe2024<empty citation>** to calculate MCSEs, see our answer to the last question.

Answer:

10.3   How many simulation repetitions will be used for each condition?

Explanation: Please also indicate whether the chosen number of simulation repetitions is based on sample size calculations, on computational constraints, rules of thumb, or any other heuristic or combination of these strategies. Formulas for sample size planning in simulation studies are provided in **Siepe2024<empty citation>**. If there is a lack of knowledge on a quantity for computing the Monte Carlo standard error (MCSE) of an estimated performance measure (e.g., the variance of the estimator is needed to compute the MCSE for the bias), pilot simulations may be needed to obtain a guess for realistic/worst-case values.

Example
We will perform 10,000 repetitions per condition. We determined this number by aiming for a MCSE of 0.005 for the type I error rate and the power under the "worst-case" rejection rate of 50% in the sense that the MCSE is maximal for a given number of repetitions ($0.50 \times (1 - 0.50)/0.005^2 = 10{,}000$ repetitions).   For illustration, we also determined the required number of repetitions to achieve a MCSE of 0.005 for the bias for each of the methods. The sample size calculation requires the empirical variance of the effect estimates $S_{\hat{\theta}}^2$ for each method. Since the empirical variances of the effect estimates can vary (pun intended) across simulation conditions, we compute the sample size using the largest estimated variance across all conditions. We obtain the empirical variance estimates for each condition and method using 100 pilot simulation runs. We found that the required sample sizes would be $S_{\hat{\theta}}^2/0.005^2 = 1{,}986$ for ANCOVA, $S_{\hat{\theta}}^2/0.005^2 = 3{,}812$ for change score analysis, and $S_{\hat{\theta}}^2/0.005^2 = 1{,}996$ for post score analysis.

Answer:

10.4 How will missing values due to non-convergence or other reasons be handled?

Explanation: 'Convergence' means that a method successfully produces the outcomes of interest (e.g., an estimate, a prediction, a p-value, a sample size, etc.) that are required for estimating the performance measures. Non-convergence of some repetitions or whole conditions of simulation studies occurs regularly, e.g., for numerical reasons. It is possible to impute non-converged repetitions, exclude all non-converged repetitions or to implement mechanisms that repeat certain parts of the simulation (such as data generation or model fitting) until convergence is achieved. Further, it is important to consider at which proportion of failed repetitions a whole condition will be excluded from the analysis. See **Pawel2024<empty citation>** for more guidance on handling missing values in simulation studies.

Example
We do not expect missing values or non-convergence. If we observe any non-convergence, we exclude the non-converged cases case-wise (keeping the converged values from the other methods in the same repetition) and report the number of non-converged cases per method and condition.

Answer:

10.5 How do you plan on interpreting the performance measures? (optional)

Explanation: It can be specified what a 'relevant difference' in performance, or what 'acceptable' and 'unacceptable' levels of performance might be to avoid post-hoc interpretation of performance. Furthermore, some researchers use regression models to analyze the results of simulations and compute effect sizes for different factors, or to assess the strength of evidence for the influence of a certain factor (**Skrondal2000**; **Chipman2022**). If such an approach will be used, please provide as many details as possible on the planned analyses.

Example
We define a type I error rate larger than 5% as non-acceptable performance. Amongst methods that exhibit acceptable performance regarding the type I error rate (within the MCSE), we consider a method X as performing better than a method Y in a certain simulation condition if the lower bound for the estimated power of method X ($\widehat{\mathrm{Pow}} - \mathrm{MCSE}$) is greater than the upper bound for the estimated power of method Y ($\widehat{\mathrm{Pow}} + \mathrm{MCSE}$).

Answer:

11 Other

11.1 Which statistical software/packages do you plan to use?

Explanation: Likely, not all software used can be prespecified before conducting the simulation. However, the main packages used for model fitting are usually known in advance and can be listed here, ideally with version numbers.

> **Example**
>
> We will use the following packages of `R` version 4.3.1 (**R2020**) in their most recent versions: The `mvtnorm` package (**Genz2009**) to generate data, the `lm()` function included in the `stats` package (**R2020**) to fit the different models, the `SimDesign` package (**Chalmers2020**) to set up and run the simulation study, and the `ggplot2` package (**Wickham2016**) to create visualizations.

Answer:

11.2 Which computational environment do you plan to use?

Explanation: Please specify the operating system and its version which you intend to use. If the study is performed on multiple machines or servers, provide information for each one of them, if possible.

> **Example**
>
> We will run the simulation study on a Windows 11 machine. The complete output of `sessionInfo()` will be saved and reported in the supplementary materials.

Answer:

11.3 Which other steps will you undertake to make simulation results reproducible? (optional)

Explanation: This can include sharing the code and full or intermediate results of the simulation in an open online repository. Additionally, this may include supplemental materials or interactive data visualizations, such as a shiny application.

> **Example**
>
> We will upload the fully reproducible simulation script and a data set containing all relevant estimates, standard errors, and *p*-values for each repetition of the simulation to OSF (https://osf.io/dfgvu/) and GitHub (https://github.com/bsiepe/SimPsychReview).

Answer:

11.4   Is there anything else you want to preregister? (optional)

Explanation: For example, the answer could include the most likely obstacles in the simulation design, and the plans to overcome them, or measures that increase the trust in the preregistration date (e.g., setting the seed based on a future event), as explained in the introduction of this template.

Example
No.

Answer:

References