University of Warsaw

Faculty of Philosophy

Zofia Broszczak

Student no. 446277

The influence of the Randomized Response Method and online environment on participants' truthfulness

Bachelor's thesis

In COGNITIVE SCIENCE

Supervisor:

dr Marek Bożykowski

Faculty of Sociology

University of Warsaw

Warsaw, June 27, 2024

Abstract

This thesis evaluates the effectiveness of the Randomized Response (RR) method in online and offline group survey settings, focusing on enhancing participants' truthfulness when answering sensitive questions. The study investigates the importance of surveying environments and the application of the RR method, particularly in online settings. After reviewing various RR designs, a specific design for applying the RR method in both online and offline group environments was created and tested. The results suggest potential differences in respondent honesty between these environments, indicating that the RR method may be more effective in digital settings. However, limitations such as small sample sizes and non-random assignment of participants are acknowledged. The study concludes with a discussion on the implications of these findings and provides suggestions for future research, emphasising the need for larger sample sizes and randomized participant assignments to enhance result reliability.

Keywords

Randomized Response method, survey honesty, online surveys, offline surveys,
research methodology

Thesis domain (Socrates-Erasmus subject area codes)

11.2 Statistics

Subject classification

Statistics
Methodology of experimental research

Tytuł pracy w języku polskim

Wpływ procedury losowania odpowiedzi (Randomised Responce) i środowiska online
na prawdomówność badanych

Contents

Acknowledgements

I would like to express my deepest gratitude to those who have supported me throughout the completion of this thesis.

First, and foremost, I wish to express my appreciation to Dr Marek Bożykowski. Your guidance and feedback have been invaluable throughout the entire process. I am extremely grateful for the support and encouragement, which made this undertaking easier and more enjoyable.

I would also like to extend my thanks to the board of examiners, especially the reviewer, Prof. Jacek Haman, for taking the time to evaluate my thesis.

Finally, I am grateful to my family and close friends for their support and understanding throughout my entire academic journey.

Introduction

Surveys as a research tool

Surveys represent one of the most widespread methodologies for collecting qualitative and quantitative data on individuals and are extensively used in many areas. They serve as indispensable tools for gathering insights into human behaviour and preferences across various disciplines and industries. According to *Cambridge Dictionary* (*Survey*, n.d.), a survey is an examination of people's opinions or behaviour made by asking people questions. Surveys are often a crucial source of knowledge for decision-making and form a basis of many scientific conclusions. Surveys with questionnaires consist of questions regarding a specific topic that are designed to allow the researcher to collect the same information in the same way from each respondent without increasing the risk of bias occurrence (Allery, 2016). The objectives of a survey encompass evaluating opinions, intuitions, or beliefs, as well as capturing factual information on a specific topic from a certain group.

Respondents' honesty in survey answers

One of the biggest issues that surveys pose is the lack of honesty in participants' answers. The impact this has on the reliability of the results is even bigger when the questions asked in the survey regard a sensitive topic. While some topics are rather straightforward and the collected answers reflect the reality quite well, others may pose problems due to the lack of subjects' sense of security and willingness to provide candid responses to the questions asked. Asking the question straightforwardly might result in the individual's refusal to respond or them giving an untruthful answer to avoid judgement, embarrassment or potential consequences. This causes non-sampling bias in sample surveys (Greenberg, Abul-Ela, Simmons, & Horvitz, 1969).

The importance of the surveying environment

To increase the reliability of the survey's results, researchers came up with ways to help participants feel safer and in turn, increase the likelihood of truthful responses. This can be done by changing the surveying environment to an online one. The idea behind this change is that people may feel more ashamed when there is a person, even a stranger, witnessing or even being in the same room while they admit to having a sensitive or shameful trait. There is research that suggests that taking part in a survey in a private setting with no unwanted witnesses, the participant feels anonymous and is more prone to answer asked questions frankly (Joinson, 1999). Said research was conducted when online surveys as well as access to technology itself were not yet widespread. More recent literature suggests that for some populations there are no significant differences in honesty in answers between online and traditional pen and paper surveys (Dodou & de Winter, 2014).

Randomised Response Method

An alternative approach to addressing the issue of untruthful answers in surveys is the Randomized Response Method (RR Method), introduced by Warner in 1965. This method offers privacy and confidentiality to the responders by introducing a layer of randomness into the survey process.

Conducting research using the original Warner's design requires the interviewer to ask the respondent two dichotomous questions at the same time, one being the opposite of the other. The interviewee gives a response - "Yes" or "No" - to one of the questions picked by chance. Thanks to the randomization layer, e.g. throwing a dice or picking a card out of a deck of cards, the interviewer gets a "Yes" or "No" response not knowing which question it refers to. The interviewees' answers do not reveal anything about a specific responder but do allow us to calculate the proportion of the researched population having a certain sensitive attribute (Warner, 1965). The original design has been modified throughout the years and many different designs of the RR technique have been created (Blair et al., 2015). The literature shows that so far the effectiveness of the different designs of the RR method has been researched in an offline environment (Ibidem).

Applying the RR Method in an online environment

In this paper, I will consider the effectiveness of the RR method with participants taking part in an auditorium questionnaire in a pen-and-paper form or online with the use of mobile devices. As mentioned in previous sections, literature proves that using the RR method in surveys increases participants' truthfulness. There is mixed evidence on the effectiveness of the online environment on participants' honesty. The integration of the RR Method with the online environment may influence the respondents' sense of privacy even further and give us more insight into people's intuitions and reactions when it comes to answering sensitive questions. This offers valuable insights into the effective utilisation of this method online or offline in probing individuals' sensitive attributes, thereby yielding results that more accurately reflect the prevalence of such characteristics.

Objective of the study

The main objective of the study is to measure what effect the online and offline environments have on the effectiveness of the RR Method and whether the researchers using the method should take that effect into consideration when conducting research using the RR method and analysing data gathered from said research.

To measure said effectiveness, an experimental study was conducted researching the influence of said method and online environment on respondents' truthfulness.

Hypotheses

The study consists of four experimental groups:

a) Group being asked a straightforward question through a pen and paper questionnaire;

b) Group being asked a straightforward question through an online questionnaire;

c) Group being asked asked a straightforward question using the RR method through a pen and paper questionnaire;

d) Group being asked a straightforward question through an online questionnaire using the RR method.

There were four hypotheses formulated:

1. Is the mean of answers in the auditorium questionnaire for the group given an online questionnaire with the sensitive question asked directly the same as for the group given the pen and paper questionnaire with the sensitive question asked directly?

2. Does using the RR method influence the results when conducting the auditorium questionnaire conducted offline (using the pen-and-paper questionnaire)?
3. Does using the RR method influence the results when conducting the auditorium questionnaire online?
4. Is the effectiveness of the RR method the same when conducting the auditorium questionnaire online and offline?

Literature Review

The Randomised Response Method was originally developed by Warner in 1965 to reduce the bias that comes with asking others straightforward questions about sensitive traits or beliefs they may potentially have. Since then, there have been many different versions of the method proposed. All of the RR methods call for a randomisation device. It can be a coin, a dice, a deck of cards or something more specific like a bag with different coloured rocks inside (Blair et. al., 2015) . The participant is to use the randomisation device without showing the result to the interviewer. They then answer the question or address the statement following the rules that were laid out earlier by the interviewer. The methods are designed in a way that allows to draw conclusions about the occurrence of the sensitive characteristic in the study group but does not reveal whether individual participants have the sensitive characteristic they were asked about. In this section, I will discuss some of the most popular versions of the RR method by presenting their designs and examples of their practical usage in the research field.

Versions of the experiment

Mirrored Question Design

The Randomised Response Method was initially devised by S. L. Warner in 1965. The original design is called the Mirrored Question design or the Warner's method. The idea is that the responders' privacy is protected by randomising whether an individual is to address the sensitive statement or its' inverse. Chance, for example picking a card out of a deck of cards or rolling the dice, decides whether the respondent answers

Question A or Question B. The interviewer does not see what card the individual drew or what number is on the dice so they do not know which statement the respondent's answer relates to.

An example of a research conducted with this design used by Blair et. al. (2015) in their paper overviewing different designs of the RR method is the experiment by Gingerich (2010). The experiment measured corruption among public bureaucrats in Bolivia, Brazil and Chile. The randomisation device was a spinner with an arrow and a circle board with regions A and B. The participants were to spin the arrow and respond "Yes" or "No" to the statement correlating with the region the arrow stopped on. If the arrow landed on region A, the participant was to address statement A and if it landed on B, they were to address statement B. The statements were as follows:

A. "I have never used, not even once, the resources of my institution for the benefit of a political party."
B. "I have used, at least once, the resources of my institution for the benefit of a political party."

The participant is not to show the board after spinning to the interviewer so that only the person addressing the statement knows which statement they are addressing. This ensures the respondent's privacy.

Unrelated Question Design

Warner's original method does not work when $p = 0.5$ which prompted Greenberg et. al. (1969) to modify the method and create the Unrelated-Question Randomised Response Design. This version of the method contains a non-sensitive question and a sensitive question.

Chi I-Cheng, et al. (1972) used this method to count the number of abortions performed by married Taiwanese women in their lifetime. The questions asked were:

A. "Have you ever had an induced abortion in your lifetime?"
B. "Were you born in a year of a horse[1]"

[1] According to the Chinese calendar, the horse is one of twelve animals that appear in the Chinese zodiac. A year of a horse happens every twelve years e.g. 1954, 1966, 1978.

The randomisation device was a bag with 35 black and 15 white stones inside. Black stones represented the sensitive question (A) and white represented the unrelated question (B). During individual interviews, each participant was given instructions to remember which colour represents which question, then pick a random stone from the bag and, while not showing the stone to the interviewer or giving any other indication of which colour of the stone they got, answer the corresponding question with only a "Yes" or "No".

Knowing how many people in the population were born in a year of a horse, the researchers estimated that 28.2 percent of the researched group has at least one induced abortion while 12.7 percent of people being a part of that same group admitted to having had an abortion when asked directly.

The difference in answers in two mentioned groups allows us to draw the conclusion that the RR method and the existence of the unrelated question B makes it more tangible to the respondent that their privacy is protected so they may be more open to giving a truthful answer if the sensitive question was selected.

Forced Response Design

In this design it is randomised whether the respondent is to answer the sensitive question truthfully or is forced to respond "Yes" or "No" regardless of the truth. The idea is that the respondents that are asked to answer truthfully will feel safe enough to do so because the interviewer will not know whether their answer was forced or honest. Each individual's privacy is protected but we still can draw conclusions about the researched group because the probabilities of a respondent being forced to say "Yes" and "No" are known.

This design was used by Chaloupka (1985) to estimate the rate of permit noncompliance when it comes to collecting marine products. To do so, members of shell clubs were randomly interviewed. The randomisation device was a table with numbers from 00 to 99 from which the participants were to choose a number. If the participant's number was between 00 and 69 ($p = 0.70$), they were to answer the sensitive question. If the number was between 70 and 84, they were forced to say "Yes" regardless of the truth ($p = 0.15$). Lastly, if it was between 85 and 99, they were to say

"No" regardless of the truth (p = 0.15). The sensitive question had two versions, A and B:

A. In the last 12 months did you collect live shells in the Capricornia Section, without first obtaining the prescribed collecting permit?
B. Given that you did obtain the prescribed permit, did you in the last 12 months collect more than the specified quantity of shells?

Members were more likely to answer truthfully when they got the sensitive question because their individual data was protected as there was a 30 percent chance that they would have been forced to give a specific answer regardless of the truth. Those who truthfully answered "Yes" felt safer because the researchers could not possibly know that a specific individual was not forced to say "Yes" and was in fact guilty of breaking the law (Chaloupka, 1985).

The results were that around 7 percent of all shell club members did not obtain a permit before collecting the shells in the researched marine park and 17.5 percent did not comply with permit conditions. This research gave reliable data surrounding the usage of marine products. With said data the marine park management could make appropriate decisions like raising the frequency of reviews based on usage.

Disguised Response Design

This version of the RR method was designed by Kuk (1990). The advantage of this approach over other versions of the RR method is that instead of "Yes" or "No", the participants answer by saying a random, non-emotionally involving word for example, a name of a colour. This design addresses the problem that some people may still feel unsure and uncomfortable giving a "Yes" or "No" answer even with other RR method related measures taken to protect their privacy.

Böckenholt et. al. (2009) used this method to estimate non-compliance with conditions that provide social security insurance benefits in the Netherlands. Said non-compliance occurs when a person receiving financial help because of a disability does not comply with the disability rules under the Dutch disability act. In this research, respondents were presented with four questions:

A. “Have you been told by your physician about a reduction in your disability symp- toms without reporting this improvement to your social welfare agency?”
B. “On your last spot–check by the social welfare agency, did you pretend to be in poorer health than you actually were?”
C. “Have you noticed personally any recovery from your disability complaints without reporting it to the social welfare agency?”
D. “Have you felt for some time now to be substantially stronger and healthier and able to work more hours, without reporting any improvement to the social welfare agency?”

The randomisation happened with two stacks of cards. In both stacks there are some red cards and black cards in different proportions. In the left stack 80% of the cards were red and in the right stack 20% of the cards were red. The respondent was asked to pick one card from the left stack and one from the right stack so that they would have had the card from the left stack in their left hand and the card from the right stack in their right hand. Then, they were asked a singular question. If the answer to the question was “Yes” they were to name the colour of the card in their left hand and if the answer was “No” they were to name the colour of the card in their right hand. This way the answers given to the interviewer were not “Yes” or “No” but “Red” or “Black”. This whole procedure was repeated for all four questions.

The point of this change was to make the responders more distanced from the question and be more comfortable with answering honestly as they would view the answer as a neutral word.

The result was that 60% of the participants responded “red” to at least one question which after using the formula presented by Blair et. al. (2015) concludes that around 67% of the respondents answered “Yes” to at least one of the questions.

Methodology

Pilot

A total of 23 third-year Cognitive Science students from the University of Warsaw were selected to take part in the pilot (see Appendix 1). The pilot’s purpose was to find an appropriate question to be asked in the survey that would allow us to test the hypothesis. The sensitive characteristic that participants would be asked about had to be one that people would likely lie about but had to also occur often enough in the researched group.

Participants were asked to choose one of the following topics that they think their peers would be the least likely to talk openly about and be most likely to lie or withhold the truth about. The topics were:

1. Whether they cheat / have ever cheated on an exam at university;
2. Whether they take / have ever taken hard drugs;
3. Whether they (think that they) have any mental health issues;
4. Whether they have ever considered or tried to commit suicide;
5. Whether they have ever destroyed or stolen something important from somebody and have not admitted to it;
6. Whether they have ever cheated or are currently cheating on their partner;

For the second question the participants were asked to put these topics in order of happening most frequently to happening least frequently among students. In question 1, 39.1% (9) of people chose topic 6, 34,8% (8) of people chose topic 4, 17.4% (4) chose topic 5, 4.3% chose topic 3 and 4.3% chose topic 1. For the analysis of the answers from the second question, we only considered answers for topic 4 and topic 6 as they got the highest percentage of answers in question 1 (34.8% and 39.1% accordingly). From the analysis of the second question we gathered that the 4th issue (considering or attempting suicide) is thought to be more common than the 6th issue (having cheated or cheating on a partner) which is why we decided for the sensitive question to be “Have you ever considered or tried to commit suicide?”.

Experiments

The actual experiment was conducted on first, second and third-year Cognitive Science students from the University of Warsaw. The survey was administered as an auditorium questionnaire during university classes. The surveys were either in the form of a traditional pen-and-paper questionnaire or an online questionnaire via *Google Forms*. Although the ‘Online’ variant differs from typical online surveys that are not conducted in a group setting, it still allows us to draw conclusions about whether the mode of administration (online or offline) affects participants' truthfulness in this context.

For each year, we chose a class that was split into four groups with around 10 students each. The University of Warsaw has a self registration policy for classes and we assume that students self registering for four groups is a randomized process.

The initial plan was to ensure that each experimental group comprised an equal number of first-year and second-year students. To achieve this, we selected two classes: one from the first year and one from the second year, each containing four groups. The intention was to form each experimental group by combining one group from the first-year class with one from the second-year class, resulting in four experimental groups of approximately equal size.

Regrettably, due to technical issues related to the scheduling of classes, we were unable to administer the survey to half of the first-year students. Consequently, I opted to collect the missing data from third-year students instead. The survey was conducted by a team of three third-year students, which may have introduced bias due to participants feeling less comfortable because they were familiar with the surveyors. To minimise the potential for bias, when surveying third-year students, the survey conductors left the room after explaining the procedure. Participants in the offline - pen-and-paper - groups were instructed to fold their completed surveys and place them into an urn. It is acknowledged that this might have introduced bias, and under ideal conditions, I recommend ensuring that participants are randomly assigned to the experimental groups.

The four experimental groups were:

1. Group that is given the 'online' ('On') survey in person and asked the sensitive question directly (without using the RR Method ('W'));
2. Group that is given the 'offline' ('Off') survey on paper in person and asked the sensitive question directly (without using the RR Method ('W'));
3. Group that is given the 'offline' ('Off') survey on paper in person and asked the sensitive question using the RR Method ('RR');

4. Group that is given the ‘online’ (‘On’) survey in person and asked the sensitive question using the RR Method (‘RR’);

Groups 1 and 4 consisted of first and second year students while groups 2 and 3 consisted of second and third year students.

Informed consent form

In each group after explaining the purpose and the course of the experiment, the participants were given an informed consent form (see Appendix 2). They were to sign the form if they agreed to take part in the research, understood possible consequences of the research and their questions were either answered or they were given the opportunity to get answers to their questions at a later date after the research was finished. They were also informed that the survey was anonymous. The consent forms were then collected and the participants were instructed to complete the survey.

Surveys

Participants from Group 1 were shown a printed QR code that they were asked to scan with their phones. The code would lead them to the survey in *Google Forms.* The sensitive question - “Have you ever considered or tied to commit suicide?” was direct (see Appendix 3). The RR method was not used. Participants from Group 2 were given the surveys in a physical paper form. They were also asked the question directly (see Appendix 3). Participants in Group 3 were given physical pen and paper surveys and they were asked the sensitive question with the help of the RR method. Participants in Group 4 were shown the *Google Forms* QR code that they were asked to scan with their phones and they were asked the question indirectly with the RR method.

For all groups the first two questions were about gender as well as field and year of study and the third one was the sensitive question asked directly or indirectly with the RR method.

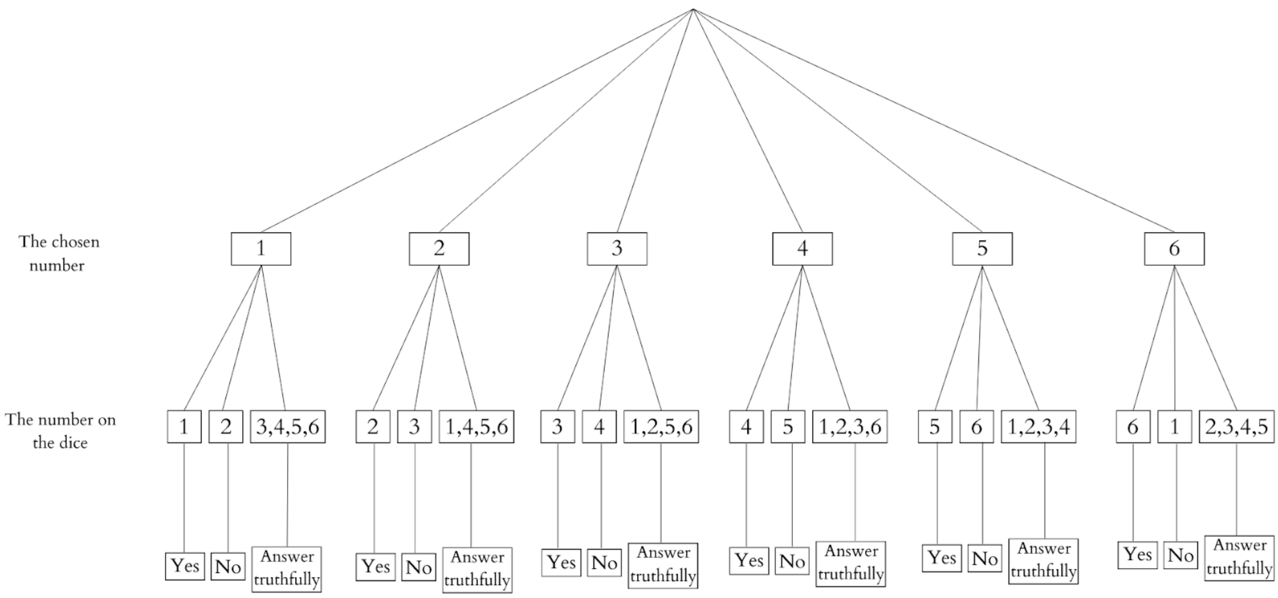
Group 3 and 4 were asked the question indirectly with a modified version of *Forced Response Design* while, again, Group 3 was given an offline paper survey and Group 4 was given an online one (see Appendix 4). They were also each given a dice as a randomisation device from the RR design.

The instructions on how to answer the third question in Groups 3 and 4 were (see Appendix 5):

”You have received a dice. Choose one number from 1 to 6 inclusive. Do not tell anyone what number you chose. After choosing the number, roll the dice.

1. If the dice lands on the number you chose, answer “Yes” regardless of the truth.
2. If the dice land on a number higher by one than the number you chose, answer “No” regardless of the truth.
    a. for number 6, the number higher by one is 1;
3. If the dice lands on a different number (than the one you chose or higher by one), answer the question truthfully.

The question: “Have you ever considered or tied to commit suicide?”.”



Instruction given to the participants in Groups 3 and 4.

We then collected the paper surveys in Group 2 and 3 and asked participants form Group 1 and 4 to submit their answers online.

Hypotheses

The four hypotheses, presented in the *Introduction*, will be verified using the formulas shown below:

While:

$E[X_i]$ - Mean of the responses (X = "Yes" = 1 or X = "No" = 0) in the 'i' group;

X_{OnRR} - Responses from the experiment conducted 'online' ('On') using the RR Method ('RR');

X_{OnW} - Responses from the experiment conducted 'online' ('On') without using the RR Method ('W');

X_{OffRR} - Responses from the experiment conducted 'offline' ('Off') using the RR Method ('RR');

X_{OffW} - Responses from the experiment conducted 'offline' ('Off') without using the RR Method ('W');

1. The mean of answers in the auditorium questionnaire for the group given an online questionnaire with the sensitive question asked directly is greater than the mean of answers for the group given the pen-and-paper questionnaire with the sensitive question asked directly.
   a. A: $E[X_{OnW}] > E[X_{OffW}]$;
2. The mean of answers after subtracting the estimated number of forced responces in the group given an auditorium pen-and-paper questionnaire with the sensitive question asked using the RR method is greater than the mean of answers in the group given the auditorium pen-and-paper questionnaire with the sensitive question asked directly.
   a. B^2: $(E[X_{OffRR}] - \frac{1}{6}) \times \frac{6}{4} > E[X_{OffW}]$;
3. The mean of answers after subtracting the estimated number of forced responses in the group given an online auditorium questionnaire with the sensitive question asked using the RR method is greater than the mean of answers in the

[2] The observed proportion of affirmative answers contains forced responses. The formula used to calculate the estimated proportion is: $p_{obs} = 1/6 + 4/6\ p_{est}$.

group given the the online auditorium questionnaire with the sensitive question asked directly.

a. C[3]: $(E[\mathrm{X}_{\mathrm{OnRR}}] - \frac{1}{6}) \times \frac{6}{4} > E[\mathrm{X}_{\mathrm{OnW}}]$;

4. The RR method is more effective when conducting the auditorium questionnaire online than when conducting it offline.

a. D: $(E[\mathrm{X}_{\mathrm{OffRR}}] - \frac{1}{6}) \times \frac{6}{4} - E[\mathrm{X}_{\mathrm{OffW}}] < (E[\mathrm{X}_{\mathrm{OnRR}}] - \frac{1}{6}) \times \frac{6}{4} - E[\mathrm{X}_{\mathrm{OnW}}]$.

Analysis

Introduction and Descriptive Statistics

The frequency distribution of collected observations is shown in Table 1.

Answer\Group	OnW	OffW	OffRR	OnRR
No	12	8	8	9
Yes	8	5	6	14
n	20	13	14	23
Proportion of “Yes” answers before subtracting forced responses	0.400	0.385	0.429	0.609

Table 1. Frequency distribution table.

In groups where the modified version of the Forced Response design of the RR Method was used, ⅙ of the participants were forced to answer “Yes” and ⅙ was forced to say “No”. For analysis purposes, we needed to subtract these records. We do not know which answers were forced nor the exact number of said forced answers so I opted to subtract exactly ⅙ of the total number of answers of answers in Groups where the RR was used from the number of “Yes” answers and from the number of “No” answers. ⅙ of the number of observations in Group 3 is 2.(3) and ⅙ of the number of observations in Group 4 is 3.8(3). Consequently, for Group 3, we subtracted 2.(3) from the number

[3] The observed proportion of affirmative answers contains forced responses. The formula used to calculate the estimated proportion is: $p_{obs} = 1/6 + 4/6\ p_{est}$.

of “Yes” answers and 2.(3) from the number of “No” answers. I repeated this process for Group 4 subtracting 3.8(3) from the number of “Yes” answers and 3.8(3) from the number of “No” answers. The final frequency distribution of the answers after subtracting the forced responses is shown in Table 2.

Answer\Group	OnW	OffW	OffRR	OnRR
No	12	8	5.667	5.167
Yes	8	5	3.667	10.167
n	20	13	9.333	15.333
Proportion of “Yes” answers after subtracting forced responses	0.400	0.385	0.393	0.663

Table 2. Frequency distribution table after subtracting the estimated number of forced responses.

Hypotheses Testing and results

To eliminate the probability that the observed differences between the means of answers in groups are of random nature, I chose to test the hypotheses while assuming the data followed a binomial probability distribution. This approach was chosen over assuming a normal distribution of the collected data due to the small sample size. The hypotheses were tested while assuming the significance level as $\alpha = 0.05$. The analysis was performed in R Studio (see Appendix 9).

All hypotheses were tested using the analysis designs outlined below.

Hypothesis 1

The first hypothesis (Hypothesis A) regards the difference between the means of answers in Groups X_{OnW} and X_{OffW} . The observed difference in means of answers between these groups is $A = E[X_{\text{OnW}}] - E[X_{\text{OffW}}] = \frac{8}{20} - \frac{5}{13} = 0.015$.

Design used for testing Hypothesis 1

The first hypothesis pertains to two groups that received the questionnaire without the RR method used. Assuming that the null hypothesis is true, we calculated a combined probability of receiving a non-forced "Yes" answer. This was done by adding the number of non-forced "Yes" answers in both groups and dividing it by the sum of the total number of observations in these two groups. The observed difference between the proportion of answers in the two groups was also calculated to be used further in the analysis.

For each of these two groups, we then created a list that consisted of all possible numbers of successes in that group. The number of successes is here understood as the number of "Yes" answers, which ranges from 0 to the total number of observations in that group. For Group "OnW" the list is *l1* = [0, 1, 2, 3, … , 18, 19, 20] and for Group "OnRR", the list is *l4* = [0, 1, 2, 3, … , 13, 14, 15], as the number of successes must be an integer. For each value on the list, the Probability mass function (PMF, Formula 1) was used to calculate the binomial probability of achieving the number of successes equal to the value, given the number of observations in the group and the combined proportion of "Yes" answers in the two groups that the hypothesis pertains to. A list containing the values calculated using PMF was created.

We then created a matrix by multiplying the lists for the two groups to obtain a matrix of probabilities for all possible combinations of successes in the two groups.

Formula 1.:

$$P_x = \binom{n}{x} p^x q^{n-x}$$

While:

P_x - binomial probability

x - number of times for a success within n trials

$\binom{n}{x}$ - number of combinations

p - probability of success on a single trial

q - probability of failure on a single trial

n - number of trials

The produced matrix was filtered to retain only the values equal or greater than the observed difference between the two groups. The resulting matrix contained only values satisfying the aforementioned condition.

By summing up the values remaining in the matrix after filtering, we obtained the probability that, assuming that the null hypothesis is true, the difference in proportions of answers between the two groups would be equal to or greater than the observed difference in true proportions of means between the groups. This probability, known as p-value, allows us to either reject or accept the null hypothesis.

An example illustrating the design for testing hypothesis 1

Let us assume that we have two experimental groups. The frequencies of answers in these groups are shown below (Table 3).

Answer\Group	Group A	Group B
No	3	2
Yes	1	3

Table 3. An example frequency distribution.

To calculate the combined probability for Groups A and B, we sum up the number of “Yes” answers and divide the sum by the total number of observations in these two groups. There are 4 “Yes” answers in total and 9 observations overall, resulting in a combined probability of 4/9 = 0.(4).

The observed difference between the means of answers is $E[X_B] - E[X_A] = 3/5 - 1/4 = 0.35$, with Group B having a higher mean of “Yes” answers than Group A.

The number of successes, percentage of successes and probability of getting each number of successes under the assumption of the combined probability = 4/9 is shown in Table 4 for Group A and in Table 5 for Group B. To calculate the probability of getting each number of successes, Formula 1 was used to calculate the PMF values for each value of all possible numbers of successes in Groups A and B.

number of successes in Group A	percentage of successes in Group A [%]	probability of the number of successes in Group A while assuming combined p = 4/9
0	0	0.095
1	25	0.305
2	50	0.366
3	75	0.195
4	100	0.039

Table 4. Table illustrating the usage of Formula 1 to calculate the probability of each number of successes occurring in Group A while assuming combined p = 4/9

number of successes in Group B	percentage of successes in Group B [%]	probability of the number of successes in Group B while assuming combined p = 4/9
0	0	0.053
1	20	0.212
2	40	0.339
3	60	0.271
4	80	0.108
5	100	0.017

Table 5. Table illustrating the usage of Formula 1. to calculate the probability of each number of successes occurring in Group B while assuming combined p = 4/9.

The number of successes in Group A and Group B are independent variables, so the probability of the occurrence of all combinations of successes in these two groups is a multiplication of probabilities from the third column in Table 4 and the third column in Table 5. By multiplying the probabilities of all possible numbers of successes in Group A and the probabilities of all possible numbers of successes in Group B, a matrix (Matrix 1) representing the probabilities of achieving each combination of successes in Groups A and B is created.

Successes Group A\Group B	0	1	2	3	4	5
0	0.0050	0.0202	0.0323	0.0258	0.0103	0.0017
1	0.0161	0.0646	0.1033	0.0826	0.0330	0.0053
2	0.0194	0.0774	0.1239	0.0991	0.0396	0.0063
3	0.0103	0.0413	0.0661	0.0528	0.0211	0.0034
4	0.0021	0.0083	0.0132	0.0106	0.0042	0.0007

Matrix 1. Matrix of probabilities of achieving each combination of successes in Group A and Group B with values rounded to 4 decimal places.

As mentioned at the beginning of this seccion, the actual difference in proportions between Group B and Group A that was observed was 0.35.

In Matrix 2 the observed differences in the proportions of success in Groups A and B are shown.

Successes Group A\Group B	0	1	2	3	4	5
0	0.00	0.20	0.40	0.60	0.80	1.00
1	-0.25	-0.05	0.15	0.35	0.55	0.75
2	-0.50	-0.30	-0.10	0.10	0.30	0.50
3	-0.75	-0.55	-0.35	-0.15	0.05	0.25
4	-1.00	-0.80	-0.60	-0.40	-0.20	0.00

Matrix 2. Matrix with the observed differences in proportions in "Yes" answers between Group B and Group A. The blue cells are the ones containing values equal to or greater than the 0.35.

A mask was created to filter the matrix and only keep values equal to or greater than said true observed difference between the two groups (Matrix 3).

Group A\Group B	0	1	2	3	4	5
0	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
1	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
2	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
3	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Matrix 3. Mask used on Matrix 1.

By applying the mask to the produced matrix (Matrix 1), we create the final matrix (Matrix 4).

Group A\Group B	0	1	2	3	4	5
0	0	0	0.0323	0.0258	0.0103	0.0017
1	0	0	0	0.0826	0.0330	0.0053
2	0	0	0	0	0	0.0063
3	0	0	0	0	0	0
4	0	0	0	0	0	0

Matrix 4. Filtered matrix with values equal to or greater than the observed difference between Groups A and B.

To calculate the p-value, we sum the values remaining in the matrix after applying the mask. As a result, we obtain *p-value* = 0.20. Since the p-value is greater than the assumed significance level ($\alpha = 0.05$) so we cannot reject the null hypothesis.

Results - Hypothesis 1

Using the design described and illustrated above, we tested the first hypothesis. Hypothesis 1 concerns the difference in means between answers in Groups "OnW" and "OffW". The hypothesis questions whether the online environment influences the results when conducting the questionnaire in a paper form and a group setting. Assuming the null hypothesis is true, the combined probability (p) for these groups -

p12 - is the sum of the number of "Yes" answers in Group "OnW" and the number of "Yes" answers in Group "OffW" divided by the sum of total observations in Group "OnW" and Group "OffW" . The combined probability is $p12 = \frac{8+5}{20+13} = 0.394$.

The observed difference in means of answers between these groups is $A = E[X_{\text{OnW}}] - E[X_{\text{OffW}}] = \frac{8}{20} - \frac{5}{13} = 0.015$. Lists with values calculated using the PMF were created, and by multiplying them, we produced the matrix of the probabilities of all combinations of successes. The matrix had 21 rows and 14 columns as Group "OnW" had 20 observations and group "OffW" had 13 observations, with the number of successes ranging from "0" to the number of observations inclusive (Appendix 6).

After filtering the matrix using the condition described above, we summed up the values in the final matrix and obtained a *p-value* of 0.448.

This does not allow us to reject the null hypothesis, indicating no statistically significant differences between Group "OnW" and Group "OffW" - when conducting an auditorium questionnaire and asking the sensitive question directly, the difference between the mean of answers for the group given an online questionnaire and the mean of answers in the group given the pen-and-paper questionnaire is not statistically significant.

Hypotheses 2 and 3

The second hypothesis (Hypothesis *B*) pertains to differences in means of answers between Groups X_{OffRR} and X_{OffW} after subtracting the calculated number of forced responses from the number of responses in Group X_{OffRR}. The difference in means of answers between these groups is 0.393 - 0.385 = 0.08. The third hypothesis (Hypothesis *C*) refers to the difference in means of answers between Groups X_{OnRR} and X_{OnW} after subtracting the calculated number of forced responses from the number of responses in Group X_{OnRR} . The difference in means of answers between these two groups is 0.663 - 0.393 = 0.27. The biggest difference was observed in regards to Hypothesis *C*.

Both the second and the third hypotheses regard the difference in means of answers between the group where the RR method was not used to ask the sensitive question (the question was asked directly) and the group where the RR method was used. Due to the use of the modified Forced Response Design of the RR method in both cases, the design for testing Hypotheses 2 and 3 was slightly different than the design used for testing Hypothesis 1. The design shown below takes into consideration the probability of different occurrences of forced responses.

Design used for testing Hypotheses 2 and 3

Let us assume that we have two experimental groups. The frequencies of answers in these groups are shown below (Table 6). Group C was asked the sensitive question directly, while Group D was asked the sensitive question with the RR method.

Answer\Group	Group C	Group D
No	3	2
Yes	1	3
n	4	5

Table 6. An example frequency distribution.

In Group D, the modified version of the Forced Response design of the RR Method was used. While the exact number of participants who were forced to give a specific answer due to the Forced Response variant is unknown, it can be estimated that ⅙ of the participants were forced to answer “Yes” and ⅙ was forced to say “No”. ⅙ of the number of observations in Group D is equal to ⅚. A fraction of ⅚ was subtracted from the number of “Yes” answers and ⅚ from the number of “No” answers in Group D. The frequency distribution of the answers after subtracting the forced responses is shown in Table 7.

Answer\Group	Group C	Group D
No	3	1⅙
Yes	1	2 ⅙
n	4	3 ⅓

Table 7. An example frequency distribution after subtracting the estimated number of forced responses from the number of answers in Group D.

The proportion of "Yes" answers observed in Group C p_{obsC} = ¼ = 0.25. The proportion of "Yes" answers observed in Group D before subtracting the forced answers is p_{obsD} = ⅗ = 0.6. After subtracting the estimated number of forced answers, the proportion of "Yes" answers is $p_{estD} = \frac{2\frac{1}{6}}{3\frac{1}{3}} = 0.65$.

To calculate the combined probability for Groups C and D, we sum up the number of "Yes" answers from Table 7 and divide the sum by the total number of observations in these two groups. The combined probability is $p_{comb} = \frac{1+2\frac{1}{6}}{4+3\frac{1}{3}} = \frac{19}{44}$. To calculate the probability of getting a "Yes" answer in Group D, we need to take the probability of getting a forced "Yes" answer into consideration. While assuming the combined probability calculated above, we get $p_{combD} = \frac{1}{6} + \frac{19}{44} \times \frac{4}{6} = \frac{5}{11}$.

According to the hypotheses, we assume that the group where the RR method is used has a greater proportion of "Yes" answers. To calculate the observed difference, we subtract the proportion of "Yes" responses in Group C from the estimated proportion of "Yes" responses among all unforced answers in Group D. The observed difference between the means of answers is $E[X_D] - E[X_C] = 0.65 - 0.25 = 0.4$.

The number of successes, probability of successes and probability of getting each number of successes under the assumption of the combined probability ($\frac{19}{44}$ for Group C and $\frac{5}{11}$ for Group D) is shown in Table 8 for Group C and in Table 9 for Group D. The probability of getting each number of successes was calculated using the binomial distribution (Formula 1).

number of successes in Group C	percentage of successes in Group C [%]	probability of the number of successes in Group C while assuming combined $p_{comb} = \frac{19}{44}$
0	0	0.104
1	25	0.317
2	50	0.361
3	75	0.183
4	100	0.034

Table 8. Table illustrating the use of binomial distribution to calculate the probability of each number of successes occurring in Group C while assuming combined $p_{comb} = \frac{19}{44}$.

In the case of the RR method, we estimate the percentage of affirmative responses among unforced answers. This percentage is then compared to the percentage of affirmative responses from the Group where the RR method is not used. Given the probability of forced responses, the estimate of the percentage is a function of the number of affirmative responses. First, we subtract the expected number of each type of response, and then we calculate the proportion of affirmative responses among the remaining unforced answers.

To illustrate the way of calculating the probability of successes, let us consider the probability of achieving two successes:

As mentioned at the beginning of this section, the estimated number of forced responses is $\frac{5}{6}$ affirmative answers and $\frac{5}{6}$ negative answers. The first step is to subtract the expected number of affirmative answers from the number of successes: $2 - \frac{5}{6} = \frac{7}{6}$. We then divide that number by the estimated number of all unforced responses, which is $5 - 2 \times \frac{5}{6}$, resulting in: $\frac{2 - \frac{5}{6}}{5 - 2 \times \frac{5}{6}} = \frac{\frac{7}{6}}{\frac{10}{3}} = \frac{7}{6} \times \frac{3}{10} = \frac{7}{20} = 0.35$.

This process was repeated for all possible numbers of successes and the percentages of successes are shown in the second column of Table 9.

Number of successes in Group D	Estimated percentage of successes in unforced answers in Group D [%]	Probability of the number of successes in Group D while assuming combined $p_{combD} = \frac{1}{6} + \frac{19}{44} \times \frac{4}{6} = \frac{5}{11}$.
0	0	0.048
1	5	0.201
2	35	0.335
3	65	0.279
4	95	0.116
5	1	0.019

Table 9. Table with percentages of successes and probabilities of each number of successes calculated with binomial distribution in Group D while assuming combined $p_{combD} = \frac{1}{6} + \frac{19}{44} \times \frac{4}{6} = \frac{5}{11}$.

The number of successes in Group C and Group D are independent variables, so the probability of the occurrence of all combinations of successes in these two groups is a multiplication of probabilities from the third column in Table 8 and the third column in Table 9. By multiplying the probabilities of all possible numbers of successes in Group C and the probabilities of all possible numbers of successes in Group D, a matrix (Matrix 5) representing the probabilities of achieving each combination of successes in Groups C and D is created..

Successes Group C\Group D	0	1	2	3	4	5
0	0.0050	0.0210	0.0349	0.0291	0.0121	0.0020
1	0.0153	0.0637	0.1062	0.0885	0.0369	0.0061
2	0.0174	0.0727	0.1211	0.1009	0.0420	0.0070
3	0.0088	0.0368	0.0614	0.0511	0.0213	0.0036

4	0.0017	0.0070	0.0117	0.0097	0.0040	0.0007

Matrix 5. Matrix of probabilities of achieving each combination of successes in Group C and Group D with values rounded to 4 decimal places.

As mentioned at the beginning of this section, the estimated difference between the proportions of unforced affirmative answers amongst all unforced answers in Group D and in Group C was 0.4.

We analyse the probability of obtaining such a combination of affirmative responses in each sample, where the estimated difference in the proportion of unforced affirmative responses is equal to or greater than the difference observed in our sample. We undertake this process to determine the statistical significance of the estimated difference in proportions of answers in Group D and Group C.

In Matrix 6, the estimated differences in the proportions of success in Groups C and D are shown.

Successes Group C\Group D	0	1	2	3	4	5
0	0.00	0.05	0.35	0.65	0.95	1.00
1	-0.25	-0.20	0.10	0.40	0.70	0.75
2	-0.50	-0.45	-0.15	0.15	0.45	0.50
3	-0.75	-0.70	-0.40	-0.10	0.20	0.25
4	-1.00	-0.95	-0.65	-0.35	-0.05	0.00

Matrix 6. Matrix with the observed differences in proportions in "Yes" answers between Group D and Group C. The blue cells are the ones containing values equal to or greater than the 0.4.

A mask was created to filter the matrix and only keep values equal to or greater than said true observed difference between the two groups (Matrix 7).

Group C\Group D	0	1	2	3	4	5
0	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
1	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
2	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
3	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Matrix 7. Mask used on Matrix 5.

By applying the mask (Matrix 7) to Matrix 5, we create the final matrix (Matrix 8).

Group C\Group D	0	1	2	3	4	5
0	0	0	0	0.0291	0.0121	0.0020
1	0	0	0	0.0885	0.0369	0.0061
2	0	0	0	0	0.0420	0.0070
3	0	0	0	0	0	0
4	0	0	0	0	0	0

Matrix 8. Filtered matrix (Matrix 5) of values equal to or greater than the observed difference between Groups C and D, only keeping the relevant entries.

To calculate the p-value, we sum the values remaining in the matrix after applying the mask. As a result, we obtain *p-value* = 0.22. Since the p-value is greater than the assumed significance level (α= 0.05), we cannot reject the null hypothesis.

Results - Hypothesis 2 and 3

Hypothesis 2 refers to the difference in proportions of "Yes" answers among unforced responses between answers in Groups "OffW" and "OffRR". The hypothesis questions whether the RR Method influences the results when conducting an auditorium questionnaire using pen-and-paper questionnaires. Assuming the null hypothesis is true, we calculated a combined p by summing the number of "Yes" answers in the "OffW"

Group and the estimated number of unforced "Yes" answers and dividing this sum by the sum of the total number of answers in the "OffW" Group and the total number of unforced answers in the "OffRR" Group. We received a combined p of 0. 3881 and an observed difference between the answers from both groups, $B = E[X_{\text{OffRR}}] - E[X_{\text{OffW}}]$, of 0.008. Then, a matrix was created by multiplying two lists of PMF values based on the combined probability for Group "OffW" and the combined probability that takes the forced responses into consideration for Group "OffRR". The matrix had 15 columns and 14 rows. This was because the number of observations in Group "OffRR" was 14 and the number of observations in Group "OffW" was 13 (see Appendix 7). After filtering the matrix and summing the remaining values we obtained a *p-value* of 0.45.

This means that if the null hypothesis is true and the *p-value* is 0.3881, the chance of drawing a pair of samples where the estimated difference in the proportions of unforced affirmative responses between Groups "OffRR" and"OffW" is greater than 0.008 (as observed empirically) is 0.45, thereby resulting in a *p-value* of 0.461. This does not allow us to reject the null hypothesis, indicating that the "OffRR" variant does not exhibit a statistically significant advantage over the "OffW" variant concerning the proportion of unforced affirmative responses.

Hypothesis 3 relates to the difference between the answers from Group "OnW" and Group "OnRR". It questions whether the RR method influences the results when conducting an auditorium questionnaire online. Assuming the null hypothesis is true, the combined p for these groups is 0.514 and the observed difference in means between the two groups, $C = E[X_{\text{OnRR}}] - E[X_{\text{OnW}}]$, is 0.263.

A matrix was created by multiplying two lists of two lists of PMF values based on the combined probability for Group "OnW" and the combined probability that takes the forced responses into consideration for Group "OnRR". The matrix had 21 rows and 24 columns because Group "OnW" had 20 observations and Group "OnRR" had 23 observations (Appendix 8). The next step was filtering the matrix to keep only the values equal to or greater than the difference between the mean of answers in the two groups. After summing the values in the filtered matrix, we found the *p-value* to be 0.08. Assuming the significance level of $\alpha = 0.05$, we could not reject the null hypothesis.

Results - Hypothesis 4

The fourth hypothesis (Hypothesis D) questioned whether the effectiveness of the RR method is the same when conducting the auditorium questionnaire online and offline. The second hypothesis testing resulted in a *p-value* of 0.45, whereas the third hypothesis testing yielded a *p-value* of 0.08. The observed difference between the means of answers between Groups "OnRR" and "OnW", denoted as $C = E[X_{\text{OnRR}}] - E[X_{\text{OnW}}]$, is 0.263, while the observed difference between the means of answers between Groups "OffRR" and "OffW", denoted as $B = E[X_{\text{OffRR}}] - E[X_{\text{OffW}}]$, is 0.008. Notably, the observed difference in the online variant exceeds that of the offline variant by 0.255.

Although neither result allowed us to reject the null hypothesis, it is noteworthy that the p-value from the third hypothesis testing is much closer to the significance level of 0.05 compared to the second hypothesis testing. This research indicates that the RR method may be more effective when conducting the auditorium questionnaire online; however, no statistically significant effects were observed. Consequently, we are unable to formally reject the null hypothesis.

Discussion

Interpretation and discussion of findings

As mentioned in the "Methodology" section, the "online" variant in my research differs from the usual online surveys as, in my research, the questionnaires are conducted in an auditorium setting. After testing the first hypothesis, I can conclude that, when asking the sensitive question directly in an auditorium questionnaire, an online environment does not affect participants' truthfulness when compared to an offline environment. Dodou and de Winter (2014) showed the hypothesis that, for some populations, an online surveying environment does not influence people's truthfulness. The results from the first hypothesis testing are consistent with the findings of Dodou and de Winter's research, indicating no significant differences in truthfulness observed in the auditorium study employing mobile devices, compared to the paper-based format.

Our research does not allow us to draw conclusions that the RR method influences the results when conducting the auditorium questionnaire conducted offline - using the pen-and-paper questionnaire. Other researchers have continuously proven the positive effect that all designs of the RR method have on participants' truthfulness. Some designs prove better than others but researchers agree that asking questions using the RR method gives results closer to reality than asking questions directly (Blair et al., 2015). The specific design used in this research was, to my knowledge, not tested by other researchers. As the survey was conducted in an auditorium setting, to ensure the participants' privacy while being in a group, we introduced another level of randomisation. It included each participant choosing a number in their mind so that even if a person sitting next to them sees what number the dice lands on, it does not reveal whether their answer was truthful or forced.

I assume that we did not manage to prove the effect of the RR method in the offline variant due to a very small sample size. It can be speculated that if the research was to be replicated on a bigger sample size, the effect would be found.

The analysis was unable to demonstrate that employing the RR method for sensitive questions in an auditorium questionnaire, where participants engage through mobile devices, has a statistically significant effect on participants' truthfulness compared to asking the sensitive question directly.

However, it is important to highlight that the result was close to being statistically significant. The observed difference in means between the two groups amounted to 26.3 percentage points.We can speculate that after replicating the research on a bigger sample size the effect would be statistically significant.

The analysis did not prove the effectiveness of the RR method offline or online so we cannot officially prove the hypothesis that the RR method used in an online environment is more effective than the RR method used in an offline setting. There was a noticeable difference between the proportion of unforced truthful answers in the "OnRR" Group and the "OffRR" Group as well as between the p-values achieved while testing the second and third hypotheses. This could lead us to speculate that the RR method used in an online environment could turn out to be more effective than the RR method used in an offline setting if the research was done on a bigger sample size.

To verify this effect I suggest replicating the research on randomly selected participants and with a much bigger sample size.

Implications of the results

The findings from this research provide valuable insights into the impact of online and offline environments on participants' truthfulness when using the RR method.

Despite the fact that the effect of the RR method in an online auditorium questionnaire could not be proven, we can speculate that the results still highlight the potential of digital tools in enhancing participant truthfulness when asked sensitive questions in surveys. The implementation of the RR method in an auditorium questionnaire in this study suggests that it can be a reliable technique for gathering truthful responses in online group settings. It can also be assumed that if large-scale random sampling studies were to confirm these findings, combining mobile tools with the RR method in auditorium surveys on sensitive issues would likely be the most effective approach. Moreover, the observed differences in the effectiveness of the RR method between online and offline settings could inform future methodological choices in survey research. Researchers might prefer online RR methods for their ease of implementation and effectiveness, especially in large-scale studies where logistical constraints make face-to-face interviews impractical.

In addition to the primary focus of this research, it was uncovered that a worryingly high percentage of participants, between 38.5% and 66.3%, reported having considered or tried to commit suicide. This alarming statistic underscores the importance of mental health awareness and intervention. It suggests that surveys, especially those employing sensitive question techniques like the RR method, can reveal critical issues beyond the scope of the original research intent.

In conclusion, this research not only contributes to the body of knowledge on the RR method and its application in different environments but also provides practical guidance for researchers and practitioners. By demonstrating the conditions under which the RR method can be effectively used, this study lays the groundwork for more robust and reliable data collection techniques in various research contexts.

Limitations of the study

Despite the effort invested in designing and executing this study, several limitations persisted.

The foremost limitation pertains to the small sample size, which undermines the reliability of the results. A larger sample size would likely offer more generalisable findings, capturing effects that might be missed when conducting research in a smaller population.

Constraints such as lack of funding and conducting research during university classes resulted in unequal group sizes, with some groups having nearly twice as many participants as others.

Another limitation the study faced was also due to class scheduling issues. The initial plan aimed for each group to comprise an approximately equal number of first- and second-year students. However, logistical constraints led to Groups "OnW" and "OnRR" consisting of first- and second-year students, while Groups "OffW" and "OffRR" included second- and third-year students. This discrepancy may have introduced bias as first- and third- year students might possess inherent differences that, in this context, could have become uncontrolled variables.

Additionally, it is important to highlight the atypical setting in which the research was conducted. Traditionally, to ensure participants' privacy, sensitive questions are asked either verbally or in writing in a face-to-face setting, with only one interviewer and one participant present at a time. However, this was not possible in this study, so the surveys were administered in an auditorium setting. This may have made the participants less comfortable that they would have been in a one-on-one setting, possibly affecting their responses.

Suggestions for future research

The limitations highlighted above underscore the need for caution when interpreting the results and highlight areas for improvement in future research designs. This subsection will focus on areas for improvement in future research and propose some expansions of my research that could broaden our understanding of the usage of the RR method and the online environment.

The most critical factors to consider in case of replicating this research are an appropriate sample size and the randomized assignment of participants to groups. Meeting these conditions will help enhance the reliability and validity of the findings.

An intriguing area for further exploration is the difference in the effectiveness of the RR technique in an auditorium setting versus a one-on-one setting. The auditorium setting is easier to implement and less time-consuming but may have some disadvantages when compared to a one-on-one survey setting. Testing the influence of the RR method on participants' truthfulness in these conditions could provide valuable insights into the method itself and influence the design process of future research using the RR technique. It could help researchers understand the advantages and disadvantages of both online variants and could result in researchers knowing how to choose the best design for their experiments' conditions.

Another worthwhile avenue of investigation would be to examine how the RR method influences the participants' truthfulness when the participants and the interviewers know each other. It could be speculated that asking the sensitive question using the RR method could be even more effective than asking the question directly in such scenarios than in my research's scenario. This research could be particularly beneficial in specific fields where there are few eligible participants or where interviewers have a pre-existing relationship with participants due to the necessary design of the research.

Conclusion

This study contributes to the understanding of the Randomised Response (RR) method's application in different survey environments. It also highlights the potential benefits of using digital tools to enhance participants' truthfulness in surveys conducted in group settings. While the findings support the use of the RR method in online auditorium questionnaires, the research did not manage to prove the effect of the RR method in the case of pen-and-paper auditorium questionnaires. Nonetheless, existing research supports the effectiveness of the RR method, and it is likely that the lack of observed effect in this study was due to the small sample size. The experiment conducted in this research was primarily illustrative of the design created to measure the effectiveness of the RR method in online and offline environments. To enhance the reliability of the results, potential future studies should replicate this research using the design presented in this paper, while increasing the sample size and ensuring that participants are randomly selected and assigned to groups. This will help to provide more robust and generalisable findings.

Bibliography

1. Allery, L. A. (2016). Design and use questionnaires for research in medical education. *Education for Primary Care*, *27*(3), 234–238. https://doi.org/10.1080/14739879.2016.1175914
2. Blair, G., Imai, K., & Zhou, Y.-Y. (2015). Design and Analysis of the Randomized Response Technique. *Journal of the American Statistical Association,* *110*(511), 1304-1319. https://doi.org/10.1080/01621459.2014.993548
3. Böckenholt, U., Barlas, S., & Van Der Heijden, P. G. M. (2009). Do Randomized-Response Designs Eliminate Response Biases? An Empirical Study of Non-Compliance Behavior. *Journal of Applied Econometrics*, *24*(3), 377–392. https://doi.org/10.1002/jae.1052
4. *Cambridge Dictionary*. (n.d.). *Survey*. Retrieved from https://dictionary.cambridge.org/dictionary/english/survey
5. Chaloupka, M. Y. (1985). Application of the randomized response technique to marine park management: An assessment of permit compliance. *Environmental Management, 9*(4), 393–398. https://doi.org/10.1007/BF01866338
6. Chi I-Cheng, Chow, L. P., & Rider, R. V. (1972). The Randomized Response Technique as Used in the Taiwan Outcome of Pregnancy Study. *Studies in Family Planning*, *3*(11), 265–269. https://doi.org/10.2307/1965247
7. Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior, 36*, 487-495. https://doi.org/10.1016/j.chb.2014.04.005
8. Gingerich, D. W. (2010). Understanding Off-the-Books Politics: Conducting Inference on the Determinants of Sensitive Behavior with Randomized Response Surveys. *Political Analysis, 18*(4), 349–380. Advance Access publication June 24, 2010. https://doi.org/10.1093/pan/mpq010
9. Greenberg, B. G., Abul-Ela, A.-L. A., Simmons, W. R., & Horvitz, D. G. (1969). The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association, 64*(326), 520-539. https://doi.org/10.1080/01621459.1969.10500991

10. Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers, 31*, 433–438. https://doi.org/10.3758/BF03200723

11. Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika, 77*, 436-438. https://doi.org/10.1093/biomet/77.2.436

12. Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association, 60*(309), 63-69. https://doi.org/10.1080/01621459.1965.10480775

Appendix

Appendix 1.

The pilot in polish.

Pilotaż badania licencjackiego - Zofia Broszczak, Kognitywistyka

Pilotaż składa się z 3 pytań. Wypełnienie ankiety powinno zająć 1-3 minuty. Badanie jest anonimowe
Nie są zbierane adresy e-mail.

* Wskazuje wymagane pytanie

1. Płeć *

   Zaznacz tylko jedną odpowiedź.

   - Kobieta
   - Mężczyzna
   - Osoba niebinarna

2. Czy jesteś studentem/ką III. Roku Kognitywistyki? *

   Zaznacz tylko jedną odpowiedź.

   - Tak
   - Nie

3. O jakich, z poniżej wymienianych kwestii, myślisz, że twoi rowiesnicy, najmniej chętnie rozmawiają i są skłonni ukrywać prawdę?

   Zaznacz wszystkie właściwe odpowiedzi.

   - [ ] Czy ściągali / ściagają na egzaminach na studiach;
   - [ ] Czy brali / biorą twarde narkotyki;
   - [ ] Czy (podejrzewają, że) maja jakieś problemy psychiczne;
   - [ ] Czy kiedykolwiek rozważali lub próbowali popełnić samobójstwo;
   - [ ] Czy kiedyś zniszczyli / ukradli komuś coś ważnego i nie przyznali się do winy;
   - [ ] Czy kiedykolwiek zdradzili lub aktualnie zdradzają swojego partnera/kę;
   - [ ] Inne: ____________

4. Ewentualne uwagi do ankiety (Opcionalne)

Ta treść nie została utworzona ani zatwierdzona przez Google.

Formularze Google

Appendix 2.

Informed consent form in polish.

Formularz świadomej zgody badanego:

Nazwisko i imię (osoby badanej):

....

Kierunek i rok studiów:

....

Wiek:

....

1. Niniejszym oświadczam, że wyrażam zgodę na udział w badaniu licencjackim Zofii Broszczak.
2. Dostałem(am) możliwość dowiedzenia się o celu badania i sposobie jego przeprowadzenia.
3. Zdaję sobie sprawę, że w badaniu zostanie zadane pytanie dotyczące wrażliwej kwestii, co może potencjalnie wpłynąć na mój komfort i samopoczucie.
4. Zostałem(am) poinformowany(a), że mogę odmówić zgody na udział w badaniu lub cofnąć ją w każdej chwili, także podczas brania udziału w badaniu.
5. Badanie jest anonimowe, nie są zbierane dane identyfikujące uczestników.
6. Uzyskałem satysfakcjonujące mnie odpowiedzi na wszelkie pytania dotyczące badania.

....

Data i czytelny podpis

Appendix 3.

Questionnaire in polish with the sensitive question asked directly; for Groups “OnW” and “OffW”.

Badanie licencjackie; Zofia Broszczak, Kognitywistyka

Przed oddaniem możesz złożyć tę kartkę na pół, żeby nie było widać Twoich odpowiedzi.

* Wskazuje wymagane pytanie

1. Płeć *

   Zaznacz tylko jedną odpowiedź.

   - Kobieta
   - Mężczyzna
   - Inne

2. Kierunek i rok studiów *

   Zaznacz tylko jedną odpowiedź.

   - Kognitywistyka Rok 1
   - Kognitywistyka Rok 2
   - Kognitywistyka Rok 3

3. Czy kiedykolwiek rozważałeś lub próbowałeś popełnić samobójstwo? *

   Zaznacz tylko jedną odpowiedź.

   - Tak
   - Nie

Appendix 4.

Questionnaire in polish with the sensitive question asked using the RR method; for Groups “OnRR” and “OffRR”.

Badanie licencjackie; Zofia Broszczak, Kognitywistyka

Przed oddaniem możesz złożyć tę kartkę na pół, żeby nie było widać Twoich odpowiedzi.

* Wskazuje wymagane pytanie

1. Płeć *

   Zaznacz tylko jedną odpowiedź.

   - Kobieta
   - Mężczyzna
   - inne

2. Kierunek i rok studiów *

   Zaznacz tylko jedną odpowiedź.

   - Kognitywistyka Rok 1
   - Kognitywistyka Rok 2
   - Kognitywistyka Rok 3

3. Dostałeś(aś) kostkę do gry. Wybierz jedną liczbę od 1 do 6 (włącznie). Nie mów nikomu jaką liczbę wybrałeś(aś). *
   Rzuć kostką.

   1. Jeśli wypadnie numer, który obstawiłeś(aś) —> odpowiedz „Tak” niezależnie od tego, jaka jest prawda;
   2. Jeśli wypadnie numer o 1 większy niż ten, co obstawiłeś(aś) —> odpowiedz „Nie” niezależnie od tego, jaka jest prawda ;
   (Dla numeru 6, o 1 większa będzie 1)
   3. Jeśli wypadnie coś innego (niż numer, który obstawiałeś(aś) lub o 1 większy) —> odpowiedz na pytanie zgodnie z prawdą;

   Pytanie:
   Czy kiedykolwiek rozważałeś lub próbowałeś popełnić samobójstwo?

   Zaznacz tylko jedną odpowiedź.

   - Tak
   - Nie

Ta treść nie została utworzona ani zatwierdzona przez Google.

Formularze Google

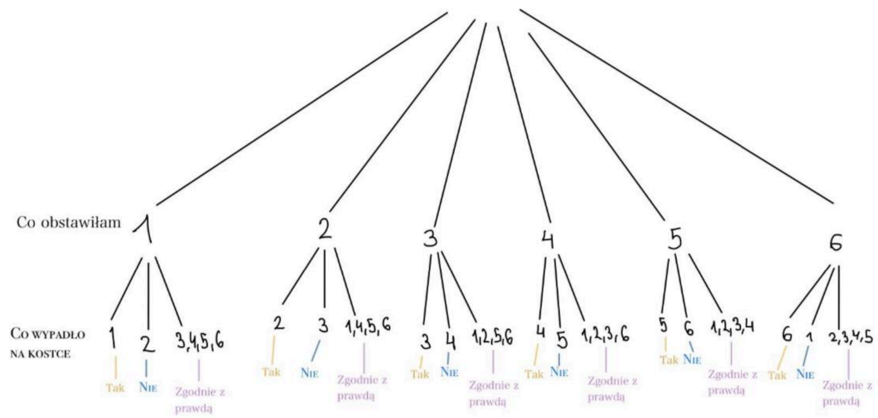
Appendix 5.

Instruction in polish given to the participants in Groups where the RR method was used ("OnRR", "OffRR").

Instrukcja dla badanych

1. Obstaw w głowie liczbę od 1 do 6 włącznie. Nie mów nikomu jaką liczbę wybrałeś.
2. Rzuć kostką.
3. Jeżeli na kostce wypadła:
    a. liczba którą obstawiłeś —> Odpowiedz „Tak" (niezależnie od tego, jaka jest prawda)
    b. liczba o jeden większa, od tej, którą obstawiłeś —> Odpowiedz „Nie" (niezależnie od tego, jaka jest prawda)
        i. dla liczby 6 liczba o jeden większa to 1!
    c. liczba inna niż te opisane w punktach a i b (czyli inna niż ta co obstawiłeś i inna niż o jeden większa od tej obstawionej) —> Odpowiedz zgodnie z prawdą.

Schemat odpowiedzi:



Przykład:

1. Wybieram liczbę 3. Zapamiętuję tę liczbę w głowie i nie mówię nikomu jaką liczbę wybrałam.
2. Rzucam kostką.
3. Jeżeli na kostce wypadła:
    a. Liczba 3 —> Zaznaczam odpowiedź „Tak".
    b. liczba 4 —> Zaznaczam odpowiedź „Nie"
    c. Liczba 1, 2, 5, 6 —> Zaznaczam odpowiedź zgodnie z prawdą o mnie. Czyli, jeśli kiedykolwiek rozważałam popełnienie samobójstwa, to zaznaczam „Tak", a jeśli nie, to zaznaczam „Nie".

Appendix 6.

Matrix of the probabilities of all combinations of numbers of successes for Hypothesis 1- blue cells are cells left after filtering the matrix.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0.000000	0.000001	0.000002	0.000005	0.000008	0.000010	0.000009	0.000006	0.000003	0.000001	0.000000	0.000000	0.000000	0.000000
1	0.000001	0.000007	0.000029	0.000068	0.000110	0.000129	0.000112	0.000073	0.000035	0.000013	0.000003	0.000001	0.000000	0.000000
2	0.000005	0.000045	0.000176	0.000419	0.000682	0.000798	0.000691	0.000449	0.000219	0.000079	0.000021	0.000004	0.000000	0.000000
3	0.000021	0.000176	0.000686	0.001636	0.002658	0.003110	0.002696	0.001752	0.000854	0.000308	0.000080	0.000014	0.000002	0.000000
4	0.000058	0.000486	0.001896	0.004519	0.007344	0.008593	0.007447	0.004840	0.002360	0.000852	0.000222	0.000039	0.000004	0.000000
5	0.000120	0.001011	0.003944	0.009400	0.015276	0.017873	0.015490	0.010068	0.004908	0.001772	0.000461	0.000082	0.000009	0.000000
6	0.000194	0.001643	0.006409	0.015276	0.024823	0.029043	0.025170	0.016361	0.007976	0.002880	0.000749	0.000133	0.000014	0.000001
7	0.000253	0.002136	0.008332	0.019858	0.032270	0.037756	0.032722	0.021269	0.010369	0.003744	0.000974	0.000173	0.000019	0.000001
8	0.000267	0.002257	0.008801	0.020975	0.034085	0.039879	0.034562	0.022465	0.010952	0.003955	0.001028	0.000182	0.000020	0.000001
9	0.000231	0.001956	0.007627	0.018179	0.029540	0.034562	0.029954	0.019470	0.009492	0.003428	0.000891	0.000158	0.000017	0.000001
10	0.000165	0.001398	0.005454	0.012998	0.021121	0.024712	0.021417	0.013921	0.006787	0.002451	0.000637	0.000113	0.000012	0.000001
11	0.000098	0.000826	0.003223	0.007680	0.012481	0.014603	0.012656	0.008226	0.004010	0.001448	0.000377	0.000067	0.000007	0.000000
12	0.000048	0.000403	0.001571	0.003744	0.006084	0.007119	0.006170	0.004010	0.001955	0.000706	0.000184	0.000033	0.000004	0.000000
13	0.000019	0.000161	0.000628	0.001498	0.002434	0.002847	0.002468	0.001604	0.000782	0.000282	0.000073	0.000013	0.000001	0.000000
14	0.000006	0.000052	0.000204	0.000487	0.000791	0.000925	0.000802	0.000521	0.000254	0.000092	0.000024	0.000004	0.000000	0.000000
15	0.000002	0.000014	0.000053	0.000127	0.000206	0.000241	0.000209	0.000136	0.000066	0.000024	0.000006	0.000001	0.000000	0.000000
16	0.000000	0.000003	0.000011	0.000026	0.000042	0.000049	0.000042	0.000028	0.000013	0.000005	0.000001	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000002	0.000004	0.000006	0.000007	0.000006	0.000004	0.000002	0.000001	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000001	0.000001	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Appendix 7.

Matrix of the probabilities of all combinations of numbers of successes for Hypothesis 2- blue cells are cells left after filtering the matrix.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0.000001	0.000007	0.000036	0.000107	0.000217	0.000321	0.000357	0.000302	0.000196	0.000097	0.000036	0.000010	0.000002	0.000000	0.000000
1	0.000006	0.000062	0.000297	0.000879	0.001789	0.002649	0.002942	0.002489	0.001612	0.000796	0.000294	0.000079	0.000015	0.000002	0.000000
2	0.000023	0.000235	0.001130	0.003345	0.006809	0.010080	0.011193	0.009470	0.006134	0.003027	0.001120	0.000302	0.000056	0.000006	0.000000
3	0.000053	0.000546	0.002626	0.007777	0.015832	0.023439	0.026026	0.022019	0.014262	0.007038	0.002605	0.000701	0.000130	0.000015	0.000001
4	0.000083	0.000865	0.004164	0.012329	0.025099	0.037159	0.041261	0.034908	0.022611	0.011159	0.004130	0.001112	0.000206	0.000023	0.000001
5	0.000095	0.000988	0.004753	0.014073	0.028650	0.042416	0.047098	0.039846	0.025809	0.012737	0.004714	0.001269	0.000235	0.000027	0.000001
6	0.000081	0.000835	0.004019	0.011899	0.024224	0.035864	0.039823	0.033691	0.021822	0.010770	0.003986	0.001073	0.000199	0.000023	0.000001
7	0.000051	0.000530	0.002548	0.007546	0.015362	0.022743	0.025254	0.021365	0.013839	0.006829	0.002528	0.000680	0.000126	0.000014	0.000001
8	0.000024	0.000252	0.001212	0.003589	0.007306	0.010817	0.012011	0.010161	0.006582	0.003248	0.001202	0.000324	0.000060	0.000007	0.000000
9	0.000009	0.000089	0.000427	0.001264	0.002574	0.003811	0.004231	0.003580	0.002319	0.001144	0.000424	0.000114	0.000021	0.000002	0.000000
10	0.000002	0.000023	0.000108	0.000321	0.000653	0.000967	0.001073	0.000908	0.000588	0.000290	0.000107	0.000029	0.000005	0.000001	0.000000
11	0.000000	0.000004	0.000019	0.000055	0.000113	0.000167	0.000186	0.000157	0.000102	0.000050	0.000019	0.000005	0.000001	0.000000	0.000000
12	0.000000	0.000000	0.000002	0.000006	0.000012	0.000018	0.000020	0.000017	0.000011	0.000005	0.000002	0.000001	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000001	0.000001	0.000001	0.000001	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Appendix 8.

Matrix of the probabilities of all combinations of numbers of successes for Hypothesis 3- blue cells are cells left after filtering the matrix.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000001	0.000001	0.000001	0.000002	0.000002	0.000002	0.000001	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000001	0.000003	0.000006	0.000010	0.000015	0.000018	0.000019	0.000016	0.000012	0.000008	0.000004	0.000002	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000001	0.000002	0.000007	0.000018	0.000037	0.000064	0.000093	0.000114	0.000119	0.000104	0.000077	0.000048	0.000025	0.000011	0.000004	0.000001	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000003	0.000010	0.000032	0.000080	0.000166	0.000288	0.000419	0.000514	0.000534	0.000469	0.000348	0.000217	0.000113	0.000048	0.000017	0.000005	0.000001	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000002	0.000009	0.000035	0.000108	0.000271	0.000564	0.000976	0.001418	0.001741	0.001808	0.001588	0.001178	0.000734	0.000381	0.000163	0.000056	0.000015	0.000003	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000001	0.000004	0.000023	0.000091	0.000285	0.000718	0.001491	0.002581	0.003753	0.004606	0.004783	0.004203	0.003117	0.001942	0.001009	0.000431	0.000149	0.000041	0.000008	0.000001	0.000000	0.000000
7	0.000000	0.000000	0.000001	0.000009	0.000049	0.000193	0.000603	0.001520	0.003156	0.005463	0.007943	0.009748	0.010123	0.008895	0.006598	0.004111	0.002135	0.000913	0.000316	0.000086	0.000018	0.000003	0.000000	0.000000
8	0.000000	0.000000	0.000002	0.000016	0.000084	0.000333	0.001036	0.002614	0.005428	0.009395	0.013659	0.016763	0.017408	0.015296	0.011346	0.007069	0.003671	0.001570	0.000543	0.000148	0.000031	0.000005	0.000000	0.000000
9	0.000000	0.000000	0.000003	0.000023	0.000119	0.000469	0.001462	0.003688	0.007659	0.013256	0.019272	0.023652	0.024562	0.021583	0.016009	0.009975	0.005179	0.002215	0.000767	0.000210	0.000044	0.000006	0.000001	0.000000
10	0.000000	0.000000	0.000004	0.000027	0.000138	0.000546	0.001702	0.004293	0.008916	0.015431	0.022434	0.027533	0.028592	0.025124	0.018636	0.011612	0.006029	0.002578	0.000892	0.000244	0.000051	0.000008	0.000001	0.000000
11	0.000000	0.000000	0.000004	0.000026	0.000133	0.000526	0.001638	0.004130	0.008577	0.014845	0.021583	0.026488	0.027507	0.024170	0.017929	0.011171	0.005800	0.002480	0.000859	0.000235	0.000049	0.000007	0.000001	0.000000
12	0.000000	0.000000	0.000003	0.000020	0.000106	0.000417	0.001300	0.003278	0.006808	0.011783	0.017130	0.021023	0.021832	0.019184	0.014230	0.008866	0.004604	0.001969	0.000681	0.000186	0.000039	0.000006	0.000001	0.000000
13	0.000000	0.000000	0.000002	0.000013	0.000069	0.000272	0.000846	0.002135	0.004433	0.007673	0.011156	0.013691	0.014218	0.012493	0.009267	0.005774	0.002998	0.001282	0.000444	0.000121	0.000025	0.000004	0.000000	0.000000
14	0.000000	0.000000	0.000001	0.000007	0.000036	0.000144	0.000448	0.001129	0.002346	0.004060	0.005903	0.007244	0.007523	0.006610	0.004903	0.003055	0.001586	0.000678	0.000235	0.000064	0.000013	0.000002	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000003	0.000015	0.000061	0.000190	0.000478	0.000993	0.001719	0.002499	0.003067	0.003184	0.002798	0.002076	0.001293	0.000671	0.000287	0.000099	0.000027	0.000006	0.000001	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000001	0.000005	0.000020	0.000063	0.000158	0.000328	0.000568	0.000826	0.001014	0.001053	0.000925	0.000686	0.000428	0.000222	0.000095	0.000033	0.000009	0.000002	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000001	0.000005	0.000016	0.000039	0.000082	0.000142	0.000206	0.000253	0.000262	0.000230	0.000171	0.000106	0.000055	0.000024	0.000008	0.000002	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000001	0.000003	0.000007	0.000014	0.000025	0.000036	0.000045	0.000046	0.000041	0.000030	0.000019	0.000010	0.000004	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000001	0.000002	0.000003	0.000004	0.000005	0.000005	0.000005	0.000003	0.000002	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Appendix 9.

Code in R Studio used to perform the analysis.

```
[58] library(googlesheets4)
     library(dplyr)
     library(tidyverse)
```

```
[59]
     # Authenticate Google Sheets (this will prompt you to authenticate)
     gs4_auth()
     sheet_url <- 'https://docs.google.com/spreadsheets/d/1-esORqcQagHMnPjT1VlHSjNeAMHNG0bIi4EX1n6EuNU/edit'
```

```
[3]  sheet1 <- read_sheet(sheet_url, sheet = 1)
     sheet2 <- read_sheet(sheet_url, sheet = 2)
     sheet3 <- read_sheet(sheet_url, sheet = 3)
     sheet4 <- read_sheet(sheet_url, sheet = 4)
```

```
✓ Reading from "Badanie Licencjackie Zofia Broszczak (Odpowiedzi)".

✓ Range ''Online bezpośrednie''.

✓ Reading from "Badanie Licencjackie Zofia Broszczak (Odpowiedzi)".

✓ Range ''Offline bezpośrednie''.

✓ Reading from "Badanie Licencjackie Zofia Broszczak (Odpowiedzi)".

✓ Range ''Offline RR''.

✓ Reading from "Badanie Licencjackie Zofia Broszczak (Odpowiedzi)".

✓ Range ''Online RR''.
```

```
[4]  dane <- bind_rows(sheet1, sheet2, sheet3, sheet4)
```

```
[5]  freq <- table(dane[c("Odp", "Grupa")])
     freq2 <- table(dane[c("Grupa")])
```

```
[6]  freq
```

```
        Grupa
Odp      1  2  3  4
  Nie 12  8  8  9
  Tak  8  5  6 14
```

```r
freq2
```

```
Grupa
 1  2  3  4 
20 13 14 23 
```

```r
[8] freq_df <- as.data.frame.matrix(freq)
    total_counts <- as.numeric(freq2)
    normalized_freq_df <- sweep(freq_df, 2, total_counts, "/")
    normalized_freq <- as.table(as.matrix(normalized_freq_df))

    print(normalized_freq)
```

```
            1         2         3         4
Nie 0.6000000 0.6153846 0.5714286 0.3913043
Tak 0.4000000 0.3846154 0.4285714 0.6086957
```

```r
[9] total_counts <- c(20, 13, 14, 23)  # Example counts for groups 1, 2, 3, and 4

    # Step 1: Calculate 1/6 of total counts for groups 1 and 4

    subtract_1_6_group3 <- (total_counts[3])/6
    subtract_1_6_group4 <- (total_counts[4])/6

    subtract_1_6_group3
    subtract_1_6_group4
```

```
2.33333333333333
3.83333333333333
```

```r
[10] freq["Tak", "3"] <- freq["Tak", "3"] - subtract_1_6_group3
     freq["Nie", "3"] <- freq["Nie", "3"] - subtract_1_6_group3

     freq["Tak", "4"] <- freq["Tak", "4"] - subtract_1_6_group4
     freq["Nie", "4"] <- freq["Nie", "4"] - subtract_1_6_group4
     print(freq)
```

```
     Grupa
Odp           1         2         3         4
  Nie 12.000000  8.000000  5.666667  5.166667
  Tak  8.000000  5.000000  3.666667 10.166667
```

```r
[11] freq2["3"] <- freq[1,3] +freq[2,3]
     freq2["4"] <- freq[1,4] +freq[2,4]
```

```r
[12] freq_df <- as.data.frame.matrix(freq)
     total_counts <- as.numeric(freq2)
     normalized_freq_df <- sweep(freq_df, 2, total_counts, "/")
     normalized_freq <- as.table(as.matrix(normalized_freq_df))

     print(normalized_freq)
```

```
            1         2         3         4
Nie 0.6000000 0.6153846 0.6071429 0.3369565
Tak 0.4000000 0.3846154 0.3928571 0.6630435
```

```r
[13] freq2
```

```
        1         2         3         4 
20.000000 13.000000  9.333333 15.333333 
```

H1 : online bezposrednio vs offline besposrednio (GUPA 1 VS GRUPA 2).

```
p12 = (freq[2,2] + freq[2,1]) / (freq2[1] + freq2[2])
print(p12)
```

```
        1
0.3939394
```

```
[15] observed_diff_p12 = abs(freq[2,1]/freq2[1]-freq[2,2]/freq2[2])
     observed_diff_p12 = format(observed_diff_p12, scientific = FALSE)
     print(observed_diff_p12)
```

```
            1
"0.01538462"
```

```
[16] l1<- 0:20

     # Initialize an empty list to store the results
     lista_gr1_p12 <- numeric(length(l1))

     # Loop through the values and calculate dbinom
     for (i in l1) {
       lista_gr1_p12[i + 1] <- dbinom(i, size = 20, prob = p12)
     }

     print(lista_gr1_p12)
```

```
 [1] 4.470140e-05 5.811182e-04 3.588405e-03 1.399478e-02 3.866058e-02
 [6] 8.041400e-02 1.306727e-01 1.698746e-01 1.794300e-01 1.555060e-01
[11] 1.111868e-01 6.570129e-02 3.202938e-02 1.281175e-02 4.163819e-03
[16] 1.082593e-03 2.199017e-04 3.363203e-05 3.643469e-06 2.492900e-07
[21] 8.101926e-09
```

```
[17] l2 <- 0:13

     # Initialize an empty list to store the results
     lista_gr2_p12 <- numeric(length(l2))

     # Loop through the values and calculate dbinom
     for (i in l2) {
       lista_gr2_p12[i + 1] <- dbinom(i, size = 13, prob = p12)
     }

     print(lista_gr2_p12)
```

```
 [1] 1.488363e-03 1.257666e-02 4.904899e-02 1.169001e-01 1.899626e-01
 [6] 2.222563e-01 1.926221e-01 1.252044e-01 6.103713e-02 2.204119e-02
[11] 5.730709e-03 1.015898e-03 1.100557e-04 5.502783e-06
```

```r
# Create the matrix using outer() function
result_matrix_p12 <- outer(lista_gr1_p12, lista_gr2_p12, FUN = "*")
result_matrix_p12 <- round(result_matrix_p12, digits = 6)
result_matrix_p12 <- format(result_matrix_p12, scientific = FALSE)

print(result_matrix_p12)
```

```
      [,1]       [,2]       [,3]       [,4]       [,5]       [,6]      
 [1,] "0.000000" "0.000001" "0.000002" "0.000005" "0.000008" "0.000010"
 [2,] "0.000001" "0.000007" "0.000029" "0.000068" "0.000110" "0.000129"
 [3,] "0.000005" "0.000045" "0.000176" "0.000419" "0.000682" "0.000798"
 [4,] "0.000021" "0.000176" "0.000686" "0.001636" "0.002658" "0.003110"
 [5,] "0.000058" "0.000486" "0.001896" "0.004519" "0.007344" "0.008593"
 [6,] "0.000120" "0.001011" "0.003944" "0.009400" "0.015276" "0.017873"
 [7,] "0.000194" "0.001643" "0.006409" "0.015276" "0.024823" "0.029043"
 [8,] "0.000253" "0.002136" "0.008332" "0.019858" "0.032270" "0.037756"
 [9,] "0.000267" "0.002257" "0.008801" "0.020975" "0.034085" "0.039879"
[10,] "0.000231" "0.001956" "0.007627" "0.018179" "0.029540" "0.034562"
[11,] "0.000165" "0.001398" "0.005454" "0.012998" "0.021121" "0.024712"
[12,] "0.000098" "0.000826" "0.003223" "0.007680" "0.012481" "0.014603"
[13,] "0.000048" "0.000403" "0.001571" "0.003744" "0.006084" "0.007119"
[14,] "0.000019" "0.000161" "0.000628" "0.001498" "0.002434" "0.002847"
[15,] "0.000006" "0.000052" "0.000204" "0.000487" "0.000791" "0.000925"
[16,] "0.000002" "0.000014" "0.000053" "0.000127" "0.000206" "0.000241"
[17,] "0.000000" "0.000003" "0.000011" "0.000026" "0.000042" "0.000049"
[18,] "0.000000" "0.000000" "0.000002" "0.000004" "0.000006" "0.000007"
[19,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000001" "0.000001"
[20,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
[21,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
      [,7]       [,8]       [,9]       [,10]      [,11]      [,12]     
 [1,] "0.000009" "0.000006" "0.000003" "0.000001" "0.000000" "0.000000"
 [2,] "0.000112" "0.000073" "0.000035" "0.000013" "0.000003" "0.000001"
 [3,] "0.000691" "0.000449" "0.000219" "0.000079" "0.000021" "0.000004"
 [4,] "0.002696" "0.001752" "0.000854" "0.000308" "0.000080" "0.000014"
 [5,] "0.007447" "0.004840" "0.002360" "0.000852" "0.000222" "0.000039"
 [6,] "0.015490" "0.010068" "0.004908" "0.001772" "0.000461" "0.000082"
 [7,] "0.025170" "0.016361" "0.007976" "0.002880" "0.000749" "0.000133"
 [8,] "0.032722" "0.021269" "0.010369" "0.003744" "0.000974" "0.000173"
 [9,] "0.034562" "0.022465" "0.010952" "0.003955" "0.001028" "0.000182"
[10,] "0.029954" "0.019470" "0.009492" "0.003428" "0.000891" "0.000158"
[11,] "0.021417" "0.013921" "0.006787" "0.002451" "0.000637" "0.000113"
[12,] "0.012656" "0.008226" "0.004010" "0.001448" "0.000377" "0.000067"
[13,] "0.006170" "0.004010" "0.001955" "0.000706" "0.000184" "0.000033"
[14,] "0.002468" "0.001604" "0.000782" "0.000282" "0.000073" "0.000013"
[15,] "0.000802" "0.000521" "0.000254" "0.000092" "0.000024" "0.000004"
[16,] "0.000209" "0.000136" "0.000066" "0.000024" "0.000006" "0.000001"
[17,] "0.000042" "0.000028" "0.000013" "0.000005" "0.000001" "0.000000"
[18,] "0.000006" "0.000004" "0.000002" "0.000001" "0.000000" "0.000000"
[19,] "0.000001" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
[20,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
[21,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
      [,13]      [,14]
```

[18]

```
      [,13]      [,14]     
 [1,] "0.000000" "0.000000"
 [2,] "0.000000" "0.000000"
 [3,] "0.000000" "0.000000"
 [4,] "0.000002" "0.000000"
 [5,] "0.000004" "0.000000"
 [6,] "0.000009" "0.000000"
 [7,] "0.000014" "0.000001"
 [8,] "0.000019" "0.000001"
 [9,] "0.000020" "0.000001"
[10,] "0.000017" "0.000001"
[11,] "0.000012" "0.000001"
[12,] "0.000007" "0.000000"
[13,] "0.000004" "0.000000"
[14,] "0.000001" "0.000000"
[15,] "0.000000" "0.000000"
[16,] "0.000000" "0.000000"
[17,] "0.000000" "0.000000"
[18,] "0.000000" "0.000000"
[19,] "0.000000" "0.000000"
[20,] "0.000000" "0.000000"
[21,] "0.000000" "0.000000"
```

[25]

```
# Create the matrix using outer() function
result_matrix_p12 <- outer(lista_gr1_p12, lista_gr2_p12, FUN = "*")
```

[26]

```
library(MASS)
```

[27]

```
write.matrix(result_matrix_p12, file="M_p12.csv")
```

```
n_rows_p12 <- nrow(result_matrix_p12)
n_cols_p12 <- ncol(result_matrix_p12)
print(n_rows_p12)
# Create the mask
mask_p12 <- outer(1:n_rows_p12, 1:n_cols_p12, FUN = function(i, j) (((i-1) / 20) - ((j-1) / 13)) > observed_diff_p12)
print(mask_p12)
# Apply the mask to get the filtered values
final_values_p12 = result_matrix_p12[mask_p12]
```

```
[60] n_rows_p12 <- nrow(result_matrix_p12)
     n_cols_p12 <- ncol(result_matrix_p12)
     # Create the mask
     mask_p12 <- outer(1:n_rows_p12, 1:n_cols_p12, FUN = function(i, j) (((i-1) / 20) - ((j-1) / 13)) > observed_diff_p12)
     # Apply the mask to get the filtered values
     final_values_p12 = result_matrix_p12[mask_p12]
```

```
sum(final_values_p12)
```

```
0.447814557766517
```

p.value = 0.45 - nie odrzucamy H0 - brak widocznych różnic między gr 1 a 2.

H2 : Offline z RR vs Ofline bezpośrednio (GRUPA 3 VS GRUPA 2).

```
[30] p23 = (freq[2,2] + freq[2,3]) / (freq2[2] + freq2[3])
     print(p23)
     observed_diff_p23 = abs(freq[2,2]/freq2[2]-freq[2,3]/freq2[3])
     print(observed_diff_p23)
```

```
        2
0.3880597
          2
0.008241758
```

```
[31] p3 = 1/6 +p23*4/6
     print(p3)
```

```
        2
0.4253731
```

```
l2 <- 0:13

# Initialize an empty list to store the results
lista_gr2_p23 <- numeric(length(l2))

# Loop through the values and calculate dbinom
for (i in l2) {
  lista_gr2_p23[i + 1] <- dbinom(i, size = 13, prob = p23)
}

print(lista_gr2_p23)
```

```
[32]
 [1] 1.687399e-03 1.391075e-02 5.292871e-02 1.230700e-01 1.951110e-01
 [6] 2.227120e-01 1.883094e-01 1.194157e-01 5.679527e-02 2.000917e-02
[11] 5.075498e-03 8.778023e-04 9.277585e-05 4.525651e-06
```

```
[33] l3<- 0:14

# Initialize an empty list to store the results
lista_gr3_p23 <- numeric(length(l3))

# Loop through the values and calculate dbinom
for (i in l3) {
  lista_gr3_p23[i + 1] <- dbinom(i, size = 14, prob = p3)
}

print(lista_gr3_p23)
```

```
 [1] 4.279598e-04 4.435220e-03 2.134089e-02 6.319122e-02 1.286393e-01
 [6] 1.904529e-01 2.114770e-01 1.789119e-01 1.158861e-01 5.719054e-02
[11] 2.116793e-02 5.698096e-03 1.054518e-03 1.200949e-04 6.350103e-06
```

```
result_matrix_p23 <- round(result_matrix_p23, digits = 6)
result_matrix_p23 <- format(result_matrix_p23, scientific = FALSE)
print(result_matrix_p23)
write.matrix(result_matrix_p23, file="M_p23.csv")
```

```
      [,1]       [,2]       [,3]       [,4]       [,5]       [,6]
 [1,] "0.000001" "0.000007" "0.000036" "0.000107" "0.000217" "0.000321"
 [2,] "0.000006" "0.000062" "0.000297" "0.000879" "0.001789" "0.002649"
 [3,] "0.000023" "0.000235" "0.001130" "0.003345" "0.006809" "0.010080"
 [4,] "0.000053" "0.000546" "0.002626" "0.007777" "0.015832" "0.023439"
 [5,] "0.000083" "0.000865" "0.004164" "0.012329" "0.025099" "0.037159"
 [6,] "0.000095" "0.000988" "0.004753" "0.014073" "0.028650" "0.042416"
 [7,] "0.000081" "0.000835" "0.004019" "0.011899" "0.024224" "0.035864"
 [8,] "0.000051" "0.000530" "0.002548" "0.007546" "0.015362" "0.022743"
 [9,] "0.000024" "0.000252" "0.001212" "0.003589" "0.007306" "0.010817"
[10,] "0.000009" "0.000089" "0.000427" "0.001264" "0.002574" "0.003811"
[11,] "0.000002" "0.000023" "0.000108" "0.000321" "0.000653" "0.000967"
[12,] "0.000000" "0.000004" "0.000019" "0.000055" "0.000113" "0.000167"
[13,] "0.000000" "0.000000" "0.000002" "0.000006" "0.000012" "0.000018"
[14,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000001" "0.000001"
```

```
      [,7]       [,8]       [,9]       [,10]      [,11]      [,12]
 [1,] "0.000357" "0.000302" "0.000196" "0.000097" "0.000036" "0.000010"
 [2,] "0.002942" "0.002489" "0.001612" "0.000796" "0.000294" "0.000079"
 [3,] "0.011193" "0.009470" "0.006134" "0.003027" "0.001120" "0.000302"
 [4,] "0.026026" "0.022019" "0.014262" "0.007038" "0.002605" "0.000701"
 [5,] "0.041261" "0.034908" "0.022611" "0.011159" "0.004130" "0.001112"
 [6,] "0.047098" "0.039846" "0.025809" "0.012737" "0.004714" "0.001269"
 [7,] "0.039823" "0.033691" "0.021822" "0.010770" "0.003986" "0.001073"
 [8,] "0.025254" "0.021365" "0.013839" "0.006829" "0.002528" "0.000680"
 [9,] "0.012011" "0.010161" "0.006582" "0.003248" "0.001202" "0.000324"
[10,] "0.004231" "0.003580" "0.002319" "0.001144" "0.000424" "0.000114"
[11,] "0.001073" "0.000908" "0.000588" "0.000290" "0.000107" "0.000029"
[12,] "0.000186" "0.000157" "0.000102" "0.000050" "0.000019" "0.000005"
[13,] "0.000020" "0.000017" "0.000011" "0.000005" "0.000002" "0.000001"
[14,] "0.000001" "0.000001" "0.000001" "0.000000" "0.000000" "0.000000"
      [,13]      [,14]      [,15]
 [1,] "0.000002" "0.000000" "0.000000"
 [2,] "0.000015" "0.000002" "0.000000"
 [3,] "0.000056" "0.000006" "0.000000"
 [4,] "0.000130" "0.000015" "0.000001"
 [5,] "0.000206" "0.000023" "0.000001"
 [6,] "0.000235" "0.000027" "0.000001"
 [7,] "0.000199" "0.000023" "0.000001"
 [8,] "0.000126" "0.000014" "0.000001"
 [9,] "0.000060" "0.000007" "0.000000"
[10,] "0.000021" "0.000002" "0.000000"
[11,] "0.000005" "0.000001" "0.000000"
[12,] "0.000001" "0.000000" "0.000000"
[13,] "0.000000" "0.000000" "0.000000"
[14,] "0.000000" "0.000000" "0.000000"
```

39]
```r
# Create the matrix using outer() function
result_matrix_p23 <- outer(lista_gr2_p23, lista_gr3_p23, FUN = "*")
```

40]
```r
# Define the vector
l3total <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14)

# Define the variable x
x <- subtract_1_6_group3
print(x)

# Function to process the values of i and j
process_value <- function(value, x) {
  if (value < x) {
    return(0)
  } else {
    return(value - x)
  }
}

# Initialize an empty list to store p values
pest3_list <- vector("numeric", length = 0)
```

[40]

```
# Iterate over l3total and calculate p
for (i in l3total) {
  j <- 14 - i
  i <- process_value(i, x)
  j <- process_value(j, x)
  p <- ifelse(i + j == 0, 0, i / (i + j))
  pest3_list <- c(pest3_list, p)  # Append p to the list
}

# Print the list of p values
print(pest3_list)
```

```
[1] 2.333333
 [1] 0.00000000 0.00000000 0.00000000 0.07142857 0.17857143 0.28571429
 [7] 0.39285714 0.50000000 0.60714286 0.71428571 0.82142857 0.92857143
[13] 1.00000000 1.00000000 1.00000000
```

```
# Create a vector of numbers from 0 to 13
num_vector <- 0:13

# Create a vector of each number divided by 13
pest2_list <- num_vector / 13

# Print the result list
print(pest2_list)
print(pest2_list[2])
```

```
 [1] 0.00000000 0.07692308 0.15384615 0.23076923 0.30769231 0.38461538
 [7] 0.46153846 0.53846154 0.61538462 0.69230769 0.76923077 0.84615385
[13] 0.92307692 1.00000000
[1] 0.07692308
```

[42]

```
difference_matrix <- outer(pest2_list, pest3_list, FUN = function(p2, p3) p3 - p2)

# Print the difference matrix
print("Difference matrix (p3 - p2):")
print(difference_matrix)
```

```
[1] "Difference matrix (p3 - p2):"
             [,1]        [,2]        [,3]         [,4]        [,5]        [,6]
 [1,]  0.00000000  0.00000000  0.00000000  0.071428571  0.17857143  0.28571429
 [2,] -0.07692308 -0.07692308 -0.07692308 -0.005494505  0.10164835  0.20879121
 [3,] -0.15384615 -0.15384615 -0.15384615 -0.082417582  0.02472527  0.13186813
 [4,] -0.23076923 -0.23076923 -0.23076923 -0.159340659 -0.05219780  0.05494505
 [5,] -0.30769231 -0.30769231 -0.30769231 -0.236263736 -0.12912088 -0.02197802
 [6,] -0.38461538 -0.38461538 -0.38461538 -0.313186813 -0.20604396 -0.09890110
 [7,] -0.46153846 -0.46153846 -0.46153846 -0.390109890 -0.28296703 -0.17582418
 [8,] -0.53846154 -0.53846154 -0.53846154 -0.467032967 -0.35989011 -0.25274725
 [9,] -0.61538462 -0.61538462 -0.61538462 -0.543956044 -0.43681319 -0.32967033
[10,] -0.69230769 -0.69230769 -0.69230769 -0.620879121 -0.51373626 -0.40659341
[11,] -0.76923077 -0.76923077 -0.76923077 -0.697802198 -0.59065934 -0.48351648
[12,] -0.84615385 -0.84615385 -0.84615385 -0.774725275 -0.66758242 -0.56043956
[13,] -0.92307692 -0.92307692 -0.92307692 -0.851648352 -0.74450549 -0.63736264
[14,] -1.00000000 -1.00000000 -1.00000000 -0.928571429 -0.82142857 -0.71428571
              [,7]        [,8]         [,9]       [,10]       [,11]
 [1,]  0.392857143  0.50000000  0.607142857  0.71428571  0.82142857
 [2,]  0.315934066  0.42307692  0.530219780  0.63736264  0.74450549
 [3,]  0.239010989  0.34615385  0.453296703  0.56043956  0.66758242
 [4,]  0.162087912  0.26923077  0.376373626  0.48351648  0.59065934
 [5,]  0.085164835  0.19230769  0.299450549  0.40659341  0.51373626
 [6,]  0.008241758  0.11538462  0.222527473  0.32967033  0.43681319
 [7,] -0.068681319  0.03846154  0.145604396  0.25274725  0.35989011
 [8,] -0.145604396 -0.03846154  0.068681319  0.17582418  0.28296703
 [9,] -0.222527473 -0.11538462 -0.008241758  0.09890110  0.20604396
[10,] -0.299450549 -0.19230769 -0.085164835  0.02197802  0.12912088
[11,] -0.376373626 -0.26923077 -0.162087912 -0.05494505  0.05219780
[12,] -0.453296703 -0.34615385 -0.239010989 -0.13186813 -0.02472527
[13,] -0.530219780 -0.42307692 -0.315934066 -0.20879121 -0.10164835
[14,] -0.607142857 -0.50000000 -0.392857143 -0.28571429 -0.17857143
             [,12]      [,13]      [,14]      [,15]
 [1,]  0.928571429 1.00000000 1.00000000 1.00000000
 [2,]  0.851648352 0.92307692 0.92307692 0.92307692
 [3,]  0.774725275 0.84615385 0.84615385 0.84615385
 [4,]  0.697802198 0.76923077 0.76923077 0.76923077
 [5,]  0.620879121 0.69230769 0.69230769 0.69230769
 [6,]  0.543956044 0.61538462 0.61538462 0.61538462
 [7,]  0.467032967 0.53846154 0.53846154 0.53846154
 [8,]  0.390109890 0.46153846 0.46153846 0.46153846
 [9,]  0.313186813 0.38461538 0.38461538 0.38461538
[10,]  0.236263736 0.30769231 0.30769231 0.30769231
[11,]  0.159340659 0.23076923 0.23076923 0.23076923
[12,]  0.082417582 0.15384615 0.15384615 0.15384615
[13,]  0.005494505 0.07692308 0.07692308 0.07692308
[14,] -0.071428571 0.00000000 0.00000000 0.00000000
```

```r
# Create the mask matrix based on the condition p3 - p2 > x
mask <- difference_matrix > observed_diff_p23

# Print the mask to verify it
print("Mask matrix (p3 - p2 > observed_diff_p23):")
print(mask)
```

```
[1] "Mask matrix (p3 - p2 > observed_diff_p23):"
       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12]
 [1,] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [2,] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [3,] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [4,] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [5,] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
 [7,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
 [8,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
 [9,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
[10,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
[11,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
[12,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
[13,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[14,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
      [,13] [,14] [,15]
 [1,]  TRUE  TRUE  TRUE
 [2,]  TRUE  TRUE  TRUE
 [3,]  TRUE  TRUE  TRUE
 [4,]  TRUE  TRUE  TRUE
 [5,]  TRUE  TRUE  TRUE
 [6,]  TRUE  TRUE  TRUE
 [7,]  TRUE  TRUE  TRUE
 [8,]  TRUE  TRUE  TRUE
 [9,]  TRUE  TRUE  TRUE
[10,]  TRUE  TRUE  TRUE
[11,]  TRUE  TRUE  TRUE
[12,]  TRUE  TRUE  TRUE
[13,]  TRUE  TRUE  TRUE
[14,] FALSE FALSE FALSE
```

```r
# Apply the mask to the result matrix
filtered_values_23 <- result_matrix_p23[mask]

print("Filtered values_23:")
print(filtered_values_23)

# Sum the filtered values
sum(filtered_values_23)
```

```
[1] "Filtered values_23:"
  [1] 1.066288e-04 2.170657e-04 1.789469e-03 6.808710e-03 3.213701e-04
  [6] 2.649343e-03 1.008043e-02 2.343904e-02 3.568460e-04 2.941803e-03
 [11] 1.119320e-02 2.602647e-02 4.126148e-02 3.018957e-04 2.488798e-03
 [16] 9.469574e-03 2.201868e-02 3.490767e-02 3.984583e-02 3.369078e-02
 [21] 1.955461e-04 1.612063e-03 6.133702e-03 1.426210e-02 2.261065e-02
 [26] 2.580923e-02 2.182244e-02 1.383862e-02 9.650325e-05 7.955634e-04
 [31] 3.027022e-03 7.038440e-03 1.115850e-02 1.273702e-02 1.076952e-02
 [36] 6.829449e-03 3.248152e-03 1.144336e-03 3.571873e-05 2.944618e-04
 [41] 1.120391e-03 2.605137e-03 4.130095e-03 4.714353e-03 3.986119e-03
 [46] 2.527783e-03 1.202238e-03 4.235528e-04 1.074378e-04 9.614960e-06
 [51] 7.926479e-05 3.015929e-04 7.012647e-04 1.111761e-03 1.269035e-03
 [56] 1.073005e-03 6.804422e-04 3.236249e-04 1.140142e-04 2.892068e-05
 [61] 5.001802e-06 1.779392e-06 1.466913e-05 5.581427e-05 1.297795e-04
 [66] 2.057480e-04 2.348538e-04 1.985756e-04 1.259260e-04 5.989163e-05
 [71] 2.110003e-05 5.352203e-06 9.256582e-07 9.783379e-08 2.026480e-07
 [76] 1.670611e-06 6.356470e-06 1.478008e-05 2.343184e-05 2.674659e-05
 [81] 2.261500e-05 1.434122e-05 6.820825e-06 2.403001e-06 6.095416e-07
 [86] 1.054196e-07 1.114191e-08 1.071516e-08 8.833470e-08 3.361028e-07
 [91] 7.815073e-07 1.238975e-06 1.414245e-06 1.195784e-06 7.583020e-07
 [96] 3.606559e-07 1.270603e-07 3.222994e-08 5.574135e-09 5.891363e-10
0.461069746987377
```

p.value = 0.461 - nie odrzucamy H0 - nie ma istotnych statystycznie różnic między gr 2 a 3.

H3: online bezp vs online RR (GRUPA 1 VS GRUPA 4).

joined p for goup 1 and 4

```
[45] p14 = (freq[2,1] + freq[2,4]) / (freq2[1] + freq2[4])
     print(p14)
```

```
         1
0.5141509
```

observed difference in answers between group 1 and 4

```
[46] observed_diff_p14 = abs(freq[2,1]/freq2[1]-freq[2,4]/freq2[4])
     print(observed_diff_p14)
```

```
         1
0.2630435
```

```
[47] p4 = 1/6 +p14*4/6
     print(p4)
```

```
       1
0.509434
```

```r
l1 <- 0:20

# Initialize an empty list to store the results
lista_gr1_p14 <- numeric(length(l1))

# Loop through the values and calculate dbinom
for (i in l1) {
  lista_gr1_p14[i + 1] <- dbinom(i, size = 20, prob = p14)
}

print(lista_gr1_p14)
```

```
 [1] 5.370655e-07 1.136702e-05 1.142772e-04 7.256045e-04 3.263459e-03
 [6] 1.105140e-02 2.923793e-02 6.188223e-02 1.064164e-01 1.501539e-01
[11] 1.747908e-01 1.681571e-01 1.334645e-01 8.691637e-02 4.598973e-02
[16] 1.946750e-02 6.437977e-03 1.603060e-03 2.827404e-04 3.149586e-05
[21] 1.666528e-06
```

```r
l4 <- 0:23

# Initialize an empty list to store the results
lista_gr4_p14 <- numeric(length(l4))

# Loop through the values and calculate dbinom
for (i in l4) {
  lista_gr4_p14[i + 1] <- dbinom(i, size = 23, prob = p4)
}

print(lista_gr4_p14)
```

```
 [1] 7.692048e-08 1.837216e-06 2.098666e-05 1.525569e-04 7.921223e-04
 [6] 3.125836e-03 9.738183e-03 2.455948e-02 5.100816e-02 8.828335e-02
[11] 1.283504e-01 1.575210e-01 1.635794e-01 1.437370e-01 1.066181e-01
[16] 6.643127e-02 3.449316e-02 1.474934e-02 5.105540e-03 1.395239e-03
[21] 2.897803e-04 4.298939e-05 4.058439e-06 1.832405e-07
```

```r
# Create the matrix using outer() function
result_matrix_p14 <- outer(lista_gr1_p14, lista_gr4_p14, FUN = "*")
```

```r
result_matrix_p14 <- round(result_matrix_p14, digits = 6)
result_matrix_p14 <- format(result_matrix_p14, scientific = FALSE)
print(result_matrix_p14)
write.matrix(result_matrix_p14, file="M_p14.csv")
```

```
      [,1]       [,2]       [,3]       [,4]       [,5]       [,6]      
 [1,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
 [2,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
 [3,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
 [4,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000001" "0.000002"
 [5,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000003" "0.000010"
 [6,] "0.000000" "0.000000" "0.000000" "0.000002" "0.000009" "0.000035"
 [7,] "0.000000" "0.000000" "0.000001" "0.000004" "0.000023" "0.000091"
 [8,] "0.000000" "0.000000" "0.000001" "0.000009" "0.000049" "0.000193"
 [9,] "0.000000" "0.000000" "0.000002" "0.000016" "0.000084" "0.000333"
[10,] "0.000000" "0.000000" "0.000003" "0.000023" "0.000119" "0.000469"
[11,] "0.000000" "0.000000" "0.000004" "0.000027" "0.000138" "0.000546"
[12,] "0.000000" "0.000000" "0.000004" "0.000026" "0.000133" "0.000526"
[13,] "0.000000" "0.000000" "0.000003" "0.000020" "0.000106" "0.000417"
[14,] "0.000000" "0.000000" "0.000002" "0.000013" "0.000069" "0.000272"
[15,] "0.000000" "0.000000" "0.000001" "0.000007" "0.000036" "0.000144"
[16,] "0.000000" "0.000000" "0.000000" "0.000003" "0.000015" "0.000061"
[17,] "0.000000" "0.000000" "0.000000" "0.000001" "0.000005" "0.000020"
[18,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000001" "0.000005"
[19,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000001"
[20,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
[21,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
      [,7]       [,8]       [,9]       [,10]      [,11]      [,12]     
 [1,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
 [2,] "0.000000" "0.000000" "0.000001" "0.000001" "0.000001" "0.000002"
 [3,] "0.000001" "0.000003" "0.000006" "0.000010" "0.000015" "0.000018"
 [4,] "0.000007" "0.000018" "0.000037" "0.000064" "0.000093" "0.000114"
 [5,] "0.000032" "0.000080" "0.000166" "0.000288" "0.000419" "0.000514"
 [6,] "0.000108" "0.000271" "0.000564" "0.000976" "0.001418" "0.001741"
 [7,] "0.000285" "0.000718" "0.001491" "0.002581" "0.003753" "0.004606"
 [8,] "0.000603" "0.001520" "0.003156" "0.005463" "0.007943" "0.009748"
 [9,] "0.001036" "0.002614" "0.005428" "0.009395" "0.013659" "0.016763"
[10,] "0.001462" "0.003688" "0.007659" "0.013256" "0.019272" "0.023652"
[11,] "0.001702" "0.004293" "0.008916" "0.015431" "0.022434" "0.027533"
[12,] "0.001638" "0.004130" "0.008577" "0.014845" "0.021583" "0.026488"
[13,] "0.001300" "0.003278" "0.006808" "0.011783" "0.017130" "0.021023"
[14,] "0.000846" "0.002135" "0.004433" "0.007673" "0.011156" "0.013691"
[15,] "0.000448" "0.001129" "0.002346" "0.004060" "0.005903" "0.007244"
[16,] "0.000190" "0.000478" "0.000993" "0.001719" "0.002499" "0.003067"
[17,] "0.000063" "0.000158" "0.000328" "0.000568" "0.000826" "0.001014"
[18,] "0.000016" "0.000039" "0.000082" "0.000142" "0.000206" "0.000253"
[19,] "0.000003" "0.000007" "0.000014" "0.000025" "0.000036" "0.000045"
[20,] "0.000000" "0.000001" "0.000002" "0.000003" "0.000004" "0.000005"
[21,] "0.000000" "0.000000" "0.000000" "0.000000" "0.000000" "0.000000"
```

	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
[1,]	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[2,]	"0.000002"	"0.000002"	"0.000001"	"0.000001"	"0.000000"	"0.000000"
[3,]	"0.000019"	"0.000016"	"0.000012"	"0.000008"	"0.000004"	"0.000002"
[4,]	"0.000119"	"0.000104"	"0.000077"	"0.000048"	"0.000025"	"0.000011"
[5,]	"0.000534"	"0.000469"	"0.000348"	"0.000217"	"0.000113"	"0.000048"
[6,]	"0.001808"	"0.001588"	"0.001178"	"0.000734"	"0.000381"	"0.000163"
[7,]	"0.004783"	"0.004203"	"0.003117"	"0.001942"	"0.001009"	"0.000431"
[8,]	"0.010123"	"0.008895"	"0.006598"	"0.004111"	"0.002135"	"0.000913"
[9,]	"0.017408"	"0.015296"	"0.011346"	"0.007069"	"0.003671"	"0.001570"
[10,]	"0.024562"	"0.021583"	"0.016009"	"0.009975"	"0.005179"	"0.002215"
[11,]	"0.028592"	"0.025124"	"0.018636"	"0.011612"	"0.006029"	"0.002578"
[12,]	"0.027507"	"0.024170"	"0.017929"	"0.011171"	"0.005800"	"0.002480"
[13,]	"0.021832"	"0.019184"	"0.014230"	"0.008866"	"0.004604"	"0.001969"
[14,]	"0.014218"	"0.012493"	"0.009267"	"0.005774"	"0.002998"	"0.001282"
[15,]	"0.007523"	"0.006610"	"0.004903"	"0.003055"	"0.001586"	"0.000678"
[16,]	"0.003184"	"0.002798"	"0.002076"	"0.001293"	"0.000671"	"0.000287"
[17,]	"0.001053"	"0.000925"	"0.000686"	"0.000428"	"0.000222"	"0.000095"
[18,]	"0.000262"	"0.000230"	"0.000171"	"0.000106"	"0.000055"	"0.000024"
[19,]	"0.000046"	"0.000041"	"0.000030"	"0.000019"	"0.000010"	"0.000004"
[20,]	"0.000005"	"0.000005"	"0.000003"	"0.000002"	"0.000001"	"0.000000"
[21,]	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"

	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
[1,]	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[2,]	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[3,]	"0.000001"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[4,]	"0.000004"	"0.000001"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[5,]	"0.000017"	"0.000005"	"0.000001"	"0.000000"	"0.000000"	"0.000000"
[6,]	"0.000056"	"0.000015"	"0.000003"	"0.000000"	"0.000000"	"0.000000"
[7,]	"0.000149"	"0.000041"	"0.000008"	"0.000001"	"0.000000"	"0.000000"
[8,]	"0.000316"	"0.000086"	"0.000018"	"0.000003"	"0.000000"	"0.000000"
[9,]	"0.000543"	"0.000148"	"0.000031"	"0.000005"	"0.000000"	"0.000000"
[10,]	"0.000767"	"0.000210"	"0.000044"	"0.000006"	"0.000001"	"0.000000"
[11,]	"0.000892"	"0.000244"	"0.000051"	"0.000008"	"0.000001"	"0.000000"
[12,]	"0.000859"	"0.000235"	"0.000049"	"0.000007"	"0.000001"	"0.000000"
[13,]	"0.000681"	"0.000186"	"0.000039"	"0.000006"	"0.000001"	"0.000000"
[14,]	"0.000444"	"0.000121"	"0.000025"	"0.000004"	"0.000000"	"0.000000"
[15,]	"0.000235"	"0.000064"	"0.000013"	"0.000002"	"0.000000"	"0.000000"
[16,]	"0.000099"	"0.000027"	"0.000006"	"0.000001"	"0.000000"	"0.000000"
[17,]	"0.000033"	"0.000009"	"0.000002"	"0.000000"	"0.000000"	"0.000000"
[18,]	"0.000008"	"0.000002"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[19,]	"0.000001"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[20,]	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"
[21,]	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"	"0.000000"

```
[52] # Define the vector
     14total <- 0:23

     # Define the variable x
     x <- subtract_1_6_group4
     print(x)

     # Function to process the values of i and j
     process_value <- function(value, x) {
       if (value < x) {
         return(0)
       } else {
         return(value - x)
       }
     }

     # Initialize an empty list to store p values
     pest4_list <- vector("numeric", length = 0)

     # Iterate over 23total and calculate p
     for (i in 14total) {
       j <- 23 - i
       i <- process_value(i, x)
       j <- process_value(j, x)
       p <- ifelse(i + j == 0, 0, i / (i + j))
       pest4_list <- c(pest4_list, p)  # Append p to the list
     }

     # Print the list of p values
     print(pest4_list)
```

```
[1] 3.833333
 [1] 0.00000000 0.00000000 0.00000000 0.00000000 0.01086957 0.07608696
 [7] 0.14130435 0.20652174 0.27173913 0.33695652 0.40217391 0.46739130
[13] 0.53260870 0.59782609 0.66304348 0.72826087 0.79347826 0.85869565
[19] 0.92391304 0.98913043 1.00000000 1.00000000 1.00000000 1.00000000
```

```
# Create a vector of numbers from 0 to 20
num_vector_1 <- 0:20

# Create a vector of each number divided by 20
pest1_list <- num_vector_1 / 20

# Print the result list
print(pest1_list)
```

```
 [1] 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70
[16] 0.75 0.80 0.85 0.90 0.95 1.00
```

```
difference_matrix_14 <- outer(pest1_list, pest4_list, FUN = function(p1, p4) p4 - p1)

# Print the difference matrix
print("Difference matrix_14 (p4 - p1):")
print(difference_matrix_14)
```

```
[1] "Difference matrix_14 (p4 - p1):"
       [,1]  [,2]  [,3]  [,4]        [,5]        [,6]         [,7]         [,8]
 [1,]  0.00  0.00  0.00  0.00  0.01086957  0.07608696  0.141304348  0.206521739
 [2,] -0.05 -0.05 -0.05 -0.05 -0.03913043  0.02608696  0.091304348  0.156521739
 [3,] -0.10 -0.10 -0.10 -0.10 -0.08913043 -0.02391304  0.041304348  0.106521739
 [4,] -0.15 -0.15 -0.15 -0.15 -0.13913043 -0.07391304 -0.008695652  0.056521739
 [5,] -0.20 -0.20 -0.20 -0.20 -0.18913043 -0.12391304 -0.058695652  0.006521739
 [6,] -0.25 -0.25 -0.25 -0.25 -0.23913043 -0.17391304 -0.108695652 -0.043478261
 [7,] -0.30 -0.30 -0.30 -0.30 -0.28913043 -0.22391304 -0.158695652 -0.093478261
 [8,] -0.35 -0.35 -0.35 -0.35 -0.33913043 -0.27391304 -0.208695652 -0.143478261
 [9,] -0.40 -0.40 -0.40 -0.40 -0.38913043 -0.32391304 -0.258695652 -0.193478261
[10,] -0.45 -0.45 -0.45 -0.45 -0.43913043 -0.37391304 -0.308695652 -0.243478261
[11,] -0.50 -0.50 -0.50 -0.50 -0.48913043 -0.42391304 -0.358695652 -0.293478261
[12,] -0.55 -0.55 -0.55 -0.55 -0.53913043 -0.47391304 -0.408695652 -0.343478261
[13,] -0.60 -0.60 -0.60 -0.60 -0.58913043 -0.52391304 -0.458695652 -0.393478261
[14,] -0.65 -0.65 -0.65 -0.65 -0.63913043 -0.57391304 -0.508695652 -0.443478261
[15,] -0.70 -0.70 -0.70 -0.70 -0.68913043 -0.62391304 -0.558695652 -0.493478261
[16,] -0.75 -0.75 -0.75 -0.75 -0.73913043 -0.67391304 -0.608695652 -0.543478261
[17,] -0.80 -0.80 -0.80 -0.80 -0.78913043 -0.72391304 -0.658695652 -0.593478261
[18,] -0.85 -0.85 -0.85 -0.85 -0.83913043 -0.77391304 -0.708695652 -0.643478261
[19,] -0.90 -0.90 -0.90 -0.90 -0.88913043 -0.82391304 -0.758695652 -0.693478261
[20,] -0.95 -0.95 -0.95 -0.95 -0.93913043 -0.87391304 -0.808695652 -0.743478261
[21,] -1.00 -1.00 -1.00 -1.00 -0.98913043 -0.92391304 -0.858695652 -0.793478261
             [,9]       [,10]        [,11]      [,12]      [,13]        [,14]
 [1,]  0.27173913  0.33695652  0.402173913  0.4673913  0.5326087  0.597826087
 [2,]  0.22173913  0.28695652  0.352173913  0.4173913  0.4826087  0.547826087
 [3,]  0.17173913  0.23695652  0.302173913  0.3673913  0.4326087  0.497826087
 [4,]  0.12173913  0.18695652  0.252173913  0.3173913  0.3826087  0.447826087
 [5,]  0.07173913  0.13695652  0.202173913  0.2673913  0.3326087  0.397826087
 [6,]  0.02173913  0.08695652  0.152173913  0.2173913  0.2826087  0.347826087
 [7,] -0.02826087  0.03695652  0.102173913  0.1673913  0.2326087  0.297826087
 [8,] -0.07826087 -0.01304348  0.052173913  0.1173913  0.1826087  0.247826087
 [9,] -0.12826087 -0.06304348  0.002173913  0.0673913  0.1326087  0.197826087
[10,] -0.17826087 -0.11304348 -0.047826087  0.0173913  0.0826087  0.147826087
[11,] -0.22826087 -0.16304348 -0.097826087 -0.0326087  0.0326087  0.097826087
[12,] -0.27826087 -0.21304348 -0.147826087 -0.0826087 -0.0173913  0.047826087
[13,] -0.32826087 -0.26304348 -0.197826087 -0.1326087 -0.0673913 -0.002173913
[14,] -0.37826087 -0.31304348 -0.247826087 -0.1826087 -0.1173913 -0.052173913
[15,] -0.42826087 -0.36304348 -0.297826087 -0.2326087 -0.1673913 -0.102173913
[16,] -0.47826087 -0.41304348 -0.347826087 -0.2826087 -0.2173913 -0.152173913
[17,] -0.52826087 -0.46304348 -0.397826087 -0.3326087 -0.2673913 -0.202173913
[18,] -0.57826087 -0.51304348 -0.447826087 -0.3826087 -0.3173913 -0.252173913
[19,] -0.62826087 -0.56304348 -0.497826087 -0.4326087 -0.3673913 -0.302173913
[20,] -0.67826087 -0.61304348 -0.547826087 -0.4826087 -0.4173913 -0.352173913
[21,] -0.72826087 -0.66304348 -0.597826087 -0.5326087 -0.4673913 -0.402173913
```

```
         [,15]       [,16]        [,17]        [,18]       [,19]       [,20]
 [1,]  0.66304348  0.72826087  0.793478261  0.858695652  0.92391304  0.98913043
 [2,]  0.61304348  0.67826087  0.743478261  0.808695652  0.87391304  0.93913043
 [3,]  0.56304348  0.62826087  0.693478261  0.758695652  0.82391304  0.88913043
 [4,]  0.51304348  0.57826087  0.643478261  0.708695652  0.77391304  0.83913043
 [5,]  0.46304348  0.52826087  0.593478261  0.658695652  0.72391304  0.78913043
 [6,]  0.41304348  0.47826087  0.543478261  0.608695652  0.67391304  0.73913043
 [7,]  0.36304348  0.42826087  0.493478261  0.558695652  0.62391304  0.68913043
 [8,]  0.31304348  0.37826087  0.443478261  0.508695652  0.57391304  0.63913043
 [9,]  0.26304348  0.32826087  0.393478261  0.458695652  0.52391304  0.58913043
[10,]  0.21304348  0.27826087  0.343478261  0.408695652  0.47391304  0.53913043
[11,]  0.16304348  0.22826087  0.293478261  0.358695652  0.42391304  0.48913043
[12,]  0.11304348  0.17826087  0.243478261  0.308695652  0.37391304  0.43913043
[13,]  0.06304348  0.12826087  0.193478261  0.258695652  0.32391304  0.38913043
[14,]  0.01304348  0.07826087  0.143478261  0.208695652  0.27391304  0.33913043
[15,] -0.03695652  0.02826087  0.093478261  0.158695652  0.22391304  0.28913043
[16,] -0.08695652 -0.02173913  0.043478261  0.108695652  0.17391304  0.23913043
[17,] -0.13695652 -0.07173913 -0.006521739  0.058695652  0.12391304  0.18913043
[18,] -0.18695652 -0.12173913 -0.056521739  0.008695652  0.07391304  0.13913043
[19,] -0.23695652 -0.17173913 -0.106521739 -0.041304348  0.02391304  0.08913043
[20,] -0.28695652 -0.22173913 -0.156521739 -0.091304348 -0.02608696  0.03913043
[21,] -0.33695652 -0.27173913 -0.206521739 -0.141304348 -0.07608696 -0.01086957
      [,21] [,22] [,23] [,24]
 [1,]  1.00  1.00  1.00  1.00
 [2,]  0.95  0.95  0.95  0.95
 [3,]  0.90  0.90  0.90  0.90
 [4,]  0.85  0.85  0.85  0.85
 [5,]  0.80  0.80  0.80  0.80
 [6,]  0.75  0.75  0.75  0.75
 [7,]  0.70  0.70  0.70  0.70
 [8,]  0.65  0.65  0.65  0.65
 [9,]  0.60  0.60  0.60  0.60
[10,]  0.55  0.55  0.55  0.55
[11,]  0.50  0.50  0.50  0.50
[12,]  0.45  0.45  0.45  0.45
[13,]  0.40  0.40  0.40  0.40
[14,]  0.35  0.35  0.35  0.35
[15,]  0.30  0.30  0.30  0.30
[16,]  0.25  0.25  0.25  0.25
[17,]  0.20  0.20  0.20  0.20
[18,]  0.15  0.15  0.15  0.15
[19,]  0.10  0.10  0.10  0.10
[20,]  0.05  0.05  0.05  0.05
[21,]  0.00  0.00  0.00  0.00
```

```r
# Create the mask matrix based on the condition p4 - p1 > obs_diff
mask_14 <- difference_matrix_14 > observed_diff_p14

# Print the mask to verify it
print("Mask matrix (p4 - p1 > observed_diff_p14):")
print(mask_14)
```

```
[1] "Mask matrix (p4 - p1 > observed_diff_p14):"
       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12]
 [1,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
 [2,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
 [3,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
 [4,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
 [5,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
 [6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [7,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [8,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [9,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[10,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[11,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[12,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[13,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[14,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[15,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[16,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[17,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[18,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[19,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[20,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[21,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
      [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
 [1,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [2,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [3,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [4,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [5,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [6,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [7,] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [8,] FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [9,] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[10,] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[11,] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[12,] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[13,] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[14,] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[15,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
[16,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[17,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[18,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[19,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[20,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[21,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
# Apply the mask to the result matrix
# Apply the mask to the result matrix
filtered_values_14 <- result_matrix_p14[mask_14]

print("Filtered values_14:")
print(filtered_values_14)
```

```
[1] "Filtered values_14:"
  [1] 2.739472e-08 4.741394e-08 1.003518e-06 6.893258e-08 1.458961e-06
  [6] 1.466752e-05 8.459907e-08 1.790543e-06 1.800105e-05 1.142979e-04
 [11] 5.140632e-04 8.785288e-08 1.859411e-06 1.869339e-05 1.186940e-04
 [16] 5.338348e-04 1.807782e-03 7.719617e-08 1.633861e-06 1.642585e-05
 [21] 1.042962e-04 4.690797e-04 1.588495e-03 4.202572e-03 5.726090e-08
 [26] 1.211930e-06 1.218401e-05 7.736256e-05 3.479437e-04 1.178279e-03
 [31] 3.117293e-03 6.597765e-03 3.567794e-08 7.551254e-07 7.591576e-06
 [36] 4.820282e-05 2.167957e-04 7.341587e-04 1.942313e-03 4.110915e-03
 [41] 7.069377e-03 9.974914e-03 1.852509e-08 3.920843e-07 3.941780e-06
 [46] 2.502839e-05 1.125670e-04 3.811978e-04 1.008509e-03 2.134514e-03
 [51] 3.670638e-03 5.179282e-03 6.029087e-03 7.921360e-09 1.676560e-07
 [56] 1.685512e-06 1.070218e-05 4.813385e-05 1.630009e-04 4.312401e-04
 [61] 9.127218e-04 1.569571e-03 2.214670e-03 2.578048e-03 2.480205e-03
 [66] 2.742009e-09 5.803476e-08 5.834466e-07 3.704602e-06 1.666172e-05
 [71] 5.642337e-05 1.492754e-04 3.159422e-04 5.433132e-04 7.666167e-04
 [76] 8.924013e-04 8.585326e-04 6.814082e-04 4.437550e-04 7.493345e-10
 [81] 1.585970e-08 1.594439e-07 1.012391e-06 4.553304e-06 1.541934e-05
 [86] 4.079389e-05 8.634047e-05 1.484763e-04 2.095005e-04 2.438749e-04
 [91] 2.346192e-04 1.862148e-04 1.212691e-04 6.416665e-05 1.556310e-10
 [96] 3.293938e-09 3.311527e-08 2.102659e-07 9.456862e-07 3.202479e-06
[101] 8.472578e-06 1.793225e-05 3.083738e-05 4.351164e-05 5.065093e-05
[106] 4.872861e-05 3.867538e-05 2.518665e-05 1.332692e-05 2.308812e-11
[111] 4.886611e-10 4.912705e-09 3.119329e-08 1.402941e-07 4.750930e-07
[116] 1.256921e-06 2.660279e-06 4.574776e-06 6.455024e-06 7.514149e-06
[121] 7.228970e-06 5.737556e-06 3.736482e-06 1.977070e-06 2.179648e-12
[126] 4.613235e-11 4.637868e-10 2.944821e-09 1.324455e-08 4.485144e-08
[131] 1.186604e-07 2.511452e-07 4.318845e-07 6.093904e-07 7.093777e-07
[136] 6.824552e-07 5.416574e-07 3.527448e-07 1.866465e-07 9.841218e-14
[141] 2.082899e-12 2.094021e-11 1.329602e-10 5.979980e-10 2.025065e-09
[146] 5.357575e-09 1.133933e-08 1.949980e-08 2.751428e-08 3.202876e-08
[151] 3.081319e-08 2.445610e-08 1.592660e-08 8.427183e-09
```

[57]
```r
sum(filtered_values_14)
```

0.080334256258462

p.value = 0.08 > 0.05 --> nie odrzucamy H0

Przyklad designu dla H1

```
pA = (1+3)/(4+5)
print(pA)
obs_diff_A = abs(1/4-3/5)
print(obs_diff_A)
lA <- 0:4
print(lA)
# Initialize an empty list to store the results
lista_A <- numeric(length(lA))

# Loop through the values and calculate dbinom
for (i in lA) {
  lista_A[i + 1] <- dbinom(i, size = 4, prob = pA)
}

print(lista_A)
```

```
[1] 0.4444444
[1] 0.35
[1] 0 1 2 3 4
[1] 0.09525987 0.30483158 0.36579790 0.19509221 0.03901844
```

```
lB <- 0:5

# Initialize an empty list to store the results
lista_B <- numeric(length(lB))

# Loop through the values and calculate dbinom
for (i in lB) {
  lista_B[i + 1] <- dbinom(i, size = 5, prob = pA)
}

print(lista_B)
```

```
[1] 0.05292215 0.21168860 0.33870176 0.27096140 0.10838456 0.01734153
```

```
# Create the matrix using outer() function
m<- outer(lista_A, lista_B, FUN = "*")
print(m)
```

```
            [,1]        [,2]       [,3]       [,4]        [,5]         [,6]
[1,] 0.005041357 0.020165428 0.03226468 0.02581175 0.010324699 0.0016519519
[2,] 0.016132342 0.064529370 0.10324699 0.08259759 0.033039037 0.0052862460
[3,] 0.019358811 0.077435244 0.12389639 0.09911711 0.039646845 0.0063434952
[4,] 0.010324699 0.041298797 0.06607807 0.05286246 0.021144984 0.0033831974
[5,] 0.002064940 0.008259759 0.01321561 0.01057249 0.004228997 0.0006766395
```

```
[ ] print(rounded_matrix)
```

```
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
[1,] 0.0050 0.0202 0.0323 0.0258 0.0103 0.0017
[2,] 0.0161 0.0645 0.1032 0.0826 0.0330 0.0053
[3,] 0.0194 0.0774 0.1239 0.0991 0.0396 0.0063
[4,] 0.0103 0.0413 0.0661 0.0529 0.0211 0.0034
[5,] 0.0021 0.0083 0.0132 0.0106 0.0042 0.0007
```

```
n_rows_AB <- nrow(m)
n_cols_AB <- ncol(m)
mask_AB <- outer(1:n_rows_AB, 1:n_cols_AB, FUN = function(i, j) (((j-1) / 5) - ((i-1) / 4)))
print(mask_AB)
# Apply the mask to get the filtered values
final_values_AB <- m[mask_AB]

print(final_values_AB)
```

```
      [,1]  [,2]  [,3]  [,4]  [,5] [,6]
[1,]  0.00  0.20  0.40  0.60  0.80 1.00
[2,] -0.25 -0.05  0.15  0.35  0.55 0.75
[3,] -0.50 -0.30 -0.10  0.10  0.30 0.50
[4,] -0.75 -0.55 -0.35 -0.15  0.05 0.25
[5,] -1.00 -0.80 -0.60 -0.40 -0.20 0.00
Error in m[mask_AB]: only 0's may be mixed with negative subscripts
Traceback:
```

Kolejne kroki: Wyjaśnij błąd

```
[ ] n_rows_AB <- nrow(m)
n_cols_AB <- ncol(m)
mask_AB <- outer(1:n_rows_AB, 1:n_cols_AB, FUN = function(i, j) (((j-1) / 5) - ((i-1) / 4))  >= obs_diff_A)
print(mask_AB)
# Apply the mask to get the filtered values
final_values_AB <- m[mask_AB]

print(final_values_AB)
```

```
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] FALSE FALSE  TRUE  TRUE  TRUE  TRUE
[2,] FALSE FALSE FALSE  TRUE  TRUE  TRUE
[3,] FALSE FALSE FALSE FALSE FALSE  TRUE
[4,] FALSE FALSE FALSE FALSE FALSE FALSE
[5,] FALSE FALSE FALSE FALSE FALSE FALSE
[1] 0.032264685 0.025811748 0.082597593 0.010324699 0.033039037 0.001651952
[7] 0.005286246 0.006343495
```

```
[ ] sum(final_values_AB)
```

```
0.197319455657391
```

Przyklad designu dla H2 i H3

```
pobsC = 1/4
```

```
pestD = (13/6)/(20/6)
pestD
```

```
0.65
```

```
obs_diff_CD = pestD-pobsC
print(obs_diff_CD)
```

```
[1] 0.4
```

```
pCD = (1+13/6)/(4+20/6)
print(pCD)
```

```
[1] 0.4318182
```

```
pcombD = 1/6 + pCD*4/6
print(pcombD)
```

```
[1] 0.4545455
```

```
lC <- 0:4
print(lC)
# Initialize an empty list to store the results
lista_C <- numeric(length(lC))

# Loop through the values and calculate dbinom
for (i in lC) {
  lista_C[i + 1] <- dbinom(i, size = 4, prob = pCD)
}

print(lista_C)
```

```
[1] 0 1 2 3 4
[1] 0.10421958 0.31682753 0.36118338 0.18299958 0.03476992
```

```
[ ] lD <- 0:5

    # Initialize an empty list to store the results
    lista_D <- numeric(length(lD))

    # Loop through the values and calculate dbinom
    for (i in lD) {
      lista_D[i + 1] <- dbinom(i, size = 5, prob = pcombD)
    }

    print(lista_D)
```

```
[1] 0.04828284 0.20117851 0.33529751 0.27941460 0.11642275 0.01940379
```

```
# Create the matrix using outer() function
m<- outer(lista_C, lista_D, FUN = "*")
print(m)
```

```
            [,1]        [,2]       [,3]        [,4]       [,5]         [,6]
[1,] 0.005032018 0.020966740 0.03494457 0.029120472 0.01213353 0.0020222550
[2,] 0.015297334 0.063738890 0.10623148 0.088526236 0.03688593 0.0061476553
[3,] 0.017438960 0.072662335 0.12110389 0.100919909 0.04204996 0.0070083270
[4,] 0.008835740 0.036815583 0.06135930 0.051132754 0.02130531 0.0035508857
[5,] 0.001678791 0.006994961 0.01165827 0.009715223 0.00404801 0.0006746683
```

```
[ ] result_matrix_CD <- round(m, digits = 4)
```

```
[ ] print(result_matrix_CD)
```

```
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
[1,] 0.0050 0.0210 0.0349 0.0291 0.0121 0.0020
[2,] 0.0153 0.0637 0.1062 0.0885 0.0369 0.0061
[3,] 0.0174 0.0727 0.1211 0.1009 0.0420 0.0070
[4,] 0.0088 0.0368 0.0614 0.0511 0.0213 0.0036
[5,] 0.0017 0.0070 0.0117 0.0097 0.0040 0.0007
```

```
# Define the vector
lDtotal <- c(0, 1, 2, 3, 4, 5)

# Define the variable x
x <- 5/6
print(x)

# Function to process the values of i and j
process_value <- function(value, x) {
  if (value < x) {
    return(0)
  } else {
    return(value - x)
  }
}

# Initialize an empty list to store p values
pestD_list <- vector("numeric", length = 0)

# Iterate over l3total and calculate p
for (i in lDtotal) {
  j <- 5 - i
  i <- process_value(i, x)
  j <- process_value(j, x)
  p <- ifelse(i + j == 0, 0, i / (i + j))
  pestD_list <- c(pestD_list, p)  # Append p to the list
}

# Print the list of p values
print(pestD_list)
```

```
[1] 0.8333333
[1] 0.00 0.05 0.35 0.65 0.95 1.00
```

```
# Create a vector of numbers from 0 to 4
num_vector_C <- 0:4

# Create a vector of each number divided by 4
pestC_list <- num_vector_C / 4

# Print the result list
print(pestC_list)
print(pestC_list[2])
```

```
[1] 0.00 0.25 0.50 0.75 1.00
[1] 0.25
```

```
[ ] difference_matrix <- outer(pestC_list, pestD_list, FUN = function(pC, pD) pD - pC)

    # Print the difference matrix
    print("Difference matrix (pD - pC):")
    print(difference_matrix)
```

```
[1] "Difference matrix (pD - pC):"
      [,1]  [,2]  [,3]  [,4]  [,5] [,6]
[1,]  0.00  0.05  0.35  0.65  0.95 1.00
[2,] -0.25 -0.20  0.10  0.40  0.70 0.75
[3,] -0.50 -0.45 -0.15  0.15  0.45 0.50
[4,] -0.75 -0.70 -0.40 -0.10  0.20 0.25
[5,] -1.00 -0.95 -0.65 -0.35 -0.05 0.00
```

```
# Create the mask matrix based on the condition pD - pC > x
mask <- difference_matrix >= obs_diff_CD

# Print the mask to verify it
print("Mask matrix (pD - pC > observed_diff_pCD = 0.4):")
print(mask)
```

```
[1] "Mask matrix (pD - pC > observed_diff_pCD = 0.4):"
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] FALSE FALSE FALSE  TRUE  TRUE  TRUE
[2,] FALSE FALSE FALSE  TRUE  TRUE  TRUE
[3,] FALSE FALSE FALSE FALSE  TRUE  TRUE
[4,] FALSE FALSE FALSE FALSE FALSE FALSE
[5,] FALSE FALSE FALSE FALSE FALSE FALSE
```

```
[ ] # Apply the mask to the result matrix
    filtered_values_CD <- result_matrix_CD[mask]

    print("Filtered values_23:")
    print(filtered_values_CD)

    # Sum the filtered values
    sum(filtered_values_CD)
```

```
[1] "Filtered values_23:"
[1] 0.0291 0.0885 0.0121 0.0369 0.0420 0.0020 0.0061 0.0070
0.2237
```