Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 27 Number 2, February 2022 ISSN 1531-7714

Autocorrelation Screening: A Potentially Efficient Method for Detecting Repetitive Response Patterns in Questionnaire Data

Jaroslav Gottfried, *Masaryk University,*
Stanislav Ježek, *Masaryk University*,
Maria Králová*, Masaryk University*,
Tomáš Řiháček, *Masaryk University*

Valid data are essential for making correct theoretical and practical implications. Hence, efficient methods for detecting and excluding data with dubious validity are highly valuable in any field of science. This paper introduces the idea of applying autocorrelation analysis on self-report questionnaires with single-choice numbered, preferably Likert-type, scales in order to screen out potentially invalid data, specifically repetitive response patterns. We explain mathematical principles of autocorrelation in a simple manner and illustrate how to efficiently perform detection of invalid data and how to correctly interpret the results. We conclude that autocorrelation screening could be a valuable screening tool for assessing the quality of self-report questionnaire data. We present a summary of the method's biggest strengths and weaknesses, together with functional tools to allow for an easy execution of autocorrelation screening by researchers, and even practitioners or the broad public. Our conclusions are limited by the current absence of empirical evidence about the practical usefulness of this method.

Introduction

Data quality is of utmost importance in research because low-quality data can introduce bias into analysis, decrease power, and even lead to invalid conclusions. In self-report measures, one of the sources of measurement bias are the participants themselves, specifically their response strategies or styles. Certain circumstances can promote response strategies leading to inaccurate or invalid responses. Besides deliberate lies, desirability bias, and errors due to misunderstandings, we are specifically referring to situations where respondents are not motivated enough to provide an accurate answer, resulting in providing a subjectively sufficient answer, even if the only criterion of sufficiency is to provide any answer at all. This has been called *satisficing* (Krosnick et al., 1996), *insufficient effort responding* (Huang et al., 2015), *inattentive responding*, or *careless responding* (Kam & Meyer, 2015). As these terms are mostly interchangeable, we use the term *careless responding* throughout this paper. The key characteristic of this problematic response strategy is a low level of participant's attention and effort while providing an answer. Consequently, provided answers are related to the question only superficially, and in extreme cases, even not at all.

Careless responding might not be as rare as researchers would like to think. Johnson (2005) identified 3.5% respondents who continuously provided the same answer. Oppenheimer et al. (2009) devised Instructional Manipulation Check (IMC) method for detecting careless respondents – a single item where a seemingly normal question is preceded by a lengthy instruction that asks respondents not to actually answer the question, but to do something else instead (e.g. clicking on the item title in online questionnaires, or answering in a very specific way), as a proof they have read the instruction carefully.

Oppenheimer et al. (2009) also found that 46% respondents failed to pass this check in their Study 1 and that 35% respondents failed to do so in Study 2, which shows that a large portion of respondents did not read instructions carefully. These authors also convincingly argue that data from participants who fail the test obscure the overall results and that certain effects do not emerge when people do not pay enough attention to their task. They conclude that respondents engaging in careless responding lower statistical power of a research design (p. 871). In "Many Labs" replication project (Klein et al., 2014), about 22% participants failed IMC, on average across the labs. Maniaci and Rogge (2014) developed a scale to measure carelessness and estimated 3–9% respondents engage in highly careless responding. The prevalence of careless responding appears to be slightly higher in online questionnaires (Oppenheimer et al., 2009) and it can be argued that the context and the content of administration, as well as the sample characteristics (e.g. age), play a major role here. The wide range of detected prevalence estimates of careless responding can also be due to different approaches to detecting, measuring and defining invalid responses across the studies. In any case, available research suggests that the overall prevalence of careless response styles is not negligible. Thus, it poses a threat to questionnaire data quality and validity of subsequent conclusions, prompting (not only) psychological researchers to prevent careless responding by properly adjusting their research design and instructions, or at the very least to perform a thorough data validity check prior to analyzing the data.

Data Validity Checks

There is a relatively broad variety of methods available for the identification of inaccurate or invalid responses in questionnaires. These methods comprise but are not limited to: bogus/infrequency/IMC indicators, consistency indicators, response times, long-string analysis, self-reports, or multivariate outlier analysis, for details see Meade & Craig (2012), and Curran (2016). A general disadvantage of these methods is that some of them may not be applicable for certain research designs (e.g. recording the number of mouse-clicks for paper-pencil questionnaires, or measuring response times during large-group paper-pencil questionnaire administration) and that thorough data quality checking is usually an intricate and time-consuming task requiring a certain level of statistical skill and familiarity with the methods available. In this regard, we would like to recommend a promising tool developed by Buchanan and Scofield (2018) that combines multiple indicators of invalid responses and allows for complex and relatively easy checking of data quality.

Repetitive Response Patterns

Repetitive response patterns are a specific form of careless responding. They may consist of any series of responses that are being repeated multiple times. A repetitive response pattern on a typical 5-point Likert-type scale might look like 1-2-3-4-5-1-2-3-4-5-…, for example, but many other variations are possible. Naturally, a presence of such repetitive response patterns may elicit researcher's suspicion that the specific respondent has been careless and their answers are not valid since such clear patterns are unlikely to occur when respondents are paying attention. Repetitive response patterns can differ in answer combination, length (i.e. number of answers, before the pattern begins repeating itself), range (i.e. range of answers utilized from available scale options), and consistency (the pattern may or may not be repeated exactly). The large number of possible repetitive response patterns is probably the main reason it received little attention in data quality research and no specialized tool for their detection has been developed so far. Many of the well-known methods for checking data quality are largely unable to detect repetitive response patterns. To our best knowledge, there exists no codified method that would be sensitive to the careless responding involving repetitive response patterns, except maybe for methods requiring adding specific dedicated items to a questionnaire or recording response times. However, the addition of dedicated questionnaire items or recording response times is not always feasible.

To our knowledge, there is no dedicated theoretical research on this kind of careless responding. Nevertheless, Tourangeau's et al. (2000) model of survey responding suggests that production of a seemingly arbitrary or haphazard responses may be a viable strategy for respondents with motivation or skills too low to engage in a high-attention responding process. Moreover, available research on the cognitive processes of random number sequence generation strongly suggests that people show a tendency to introduce a structure into their answers even when they are tasked to choose numbers at random. This tendency often manifests in the preference and the avoidance of certain numbers (Treisman & Faulkner, 1987), in the tendency to count, i.e. providing an increasing or decreasing sequence of numbers, or alphabetical sorting in case of random letter generation (Baddeley et al., 1998), and also in the tendency to autocorrelate the provided numbers, specifically, a tendency to provide a number similar, but not the same, in value to the number before (Towse & Valentine, 1997; Zabelina et al., 2012). Also, the generation of truly random sequences requires considerable effort since it involves extensive use of working memory (Treisman & Faulkner, 1987). Such effort is not compatible with our premise of carelessness. Consequently, a carelessly produced sequence of responses intended to be random or to seem random actually is not random but rather haphazard in nature, involving less attention and effort and thus being more prone to the aforementioned biases. Therefore, we argue that careless respondents provide strongly autocorrelated responses.

We acknowledge that the cognitive process of the random sequence generation considerably differs from the cognitive process of carelessly answering a questionnaire. Still, we feel there might be some common underlying factors for both of them. We see the aforementioned findings as the possible theoretical grounds for the existence of repetitive response patterns in questionnaire social research.

Unfortunately, we are not aware of any empirical research that would focus on this kind of careless responding. As such, we do not know or even dare to estimate how prevalent this type of careless responding generally is and, consequently, how much serious problem it constitutes for data analysis. But, given our judgment that many existing methods are likely to not be very sensitive to repetitive response patterns due to their variety, we consider it worthwhile to develop a dedicated method for the detection of repetitive response patterns, which would allow for a better assessment of their prevalence and the seriousness of bias for research results.

After studying the problem, we are convinced that the problems with repetitive response pattern detection can be overcome, and that this type of careless responding could be detected quite reliably without the need to measure response times or add dedicated questionnaire items.

Aim of the Paper

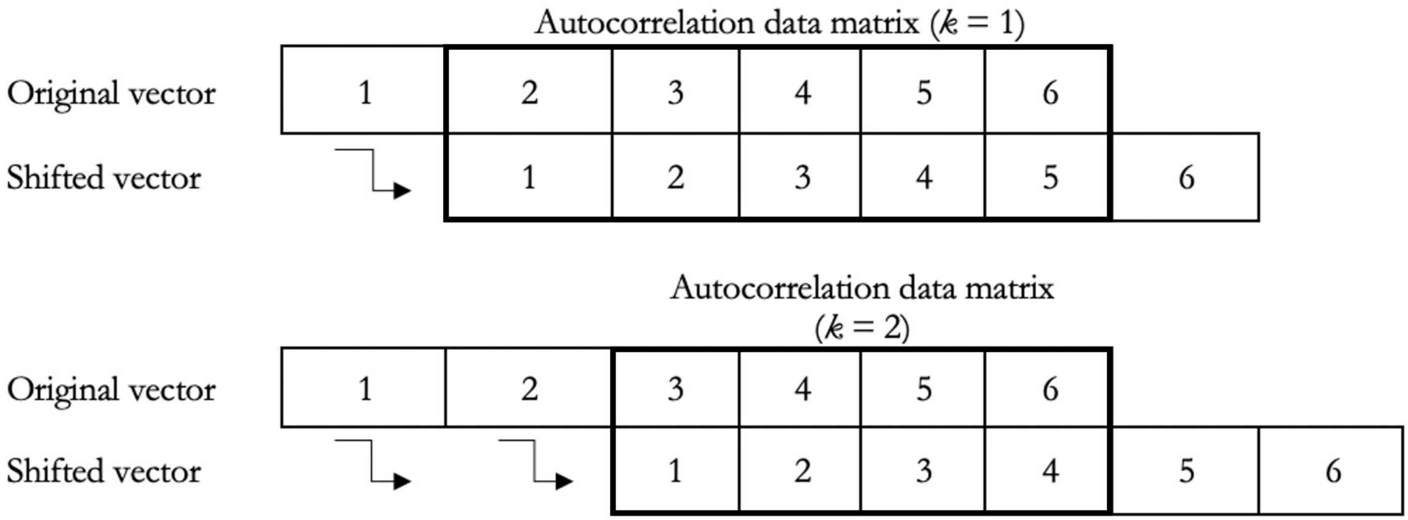
This paper aims to provide arguments that the autocorrelation screening method constitutes an efficient way to detect repetitive response patterns and thus this method could contribute to a higher quality of research data, if used in practice. In the rest of this paper, we describe fundamental principles of autocorrelation, explain how autocorrelation function could be used for screening out certain invalid questionnaire data, and present an easy-to-use tool for immediate practical application.

Introduction to Autocorrelation

Autocorrelation is the (usually Pearson's) correlation of time series data with a shifted copy of themselves. In our present case the data are an individual respondent's questionnaire answers. The intuitive term *shift* is often called *lag* in econometrics literature and is denoted by k. If $k = 1$, the data pairs over which the correlation is computed are the 1st and 2nd response, 2nd and 3rd, 3rd and 4th, and so forth. Figure 1 presents a construction process of instrumental data matrices (highlighted in bold) from which the autocorrelation $k = 1$ and $k = 2$ can be calculated, respectively. Autocorrelation is predominantly used in time series models in econometrics and computer engineering. In such models, autocorrelation serves as a means to detect and control for periodic fluctuations, so the corresponding variation is controlled for and other trends become more apparent. In questionnaire data context, autocorrelation for $k = 1$ can be described as an overall within-subject correlation of all adjacent item answers (i.e. first answer with the second one, second answer with the third one, etc.), autocorrelation for $k = 2$ can be described as a correlation of answers exactly two "positions" away (i.e. first answer with the third one, second answer with the fourth one, etc.), and so on.

Autocorrelation coefficients are easy to interpret, because they are practically identical to standard correlation coefficients, with the same possible range from -1 to 1, when -1 indicates perfect negative correlation, 0 indicates no correlation, and 1 indicates perfect positive correlation. This means, that if there

Figure 1. Two examples of instrumental data matrices (highlighted in bold) for k = 1 and k = 2 autocorrelation calculation



a perfect repeating pattern over k values the k-lag autocorrelation would be equal to $r = 1$. However, the repeating patterns are rarely perfect, necessitating some experience with the interpretation on autocorrelations as indicators of careless responding.

Autocorrelation Screening in Theory

We propose a method for autocorrelation screening that can be applied to questionnaire and survey data in order to detect some potentially invalid answers. The autocorrelation screening itself should take place during the data inspection and cleaning phase before main data analysis. The proposed screening procedure is actually very simple and can be done even with a rudimentary software. Provided the questionnaire items are accompanied by single-choice scales, preferably of at least ordinal nature or Likert-type where numbers can be reasonably assigned to choices, the only fundamental requirement for autocorrelation screening is that the order in which the items were responded must be known and data for each respondent must be arranged in this order prior to autocorrelation screening. Since this analysis is meant to detect answer patterns repeated in time, it does not produce meaningful results unless data are arranged by chronological order for each respondent. Subsequently, for each respondent, autocorrelation coefficients for all reasonable *lag* (k) values are calculated and the highest absolute autocorrelation coefficient is kept. The resulting value should indicate the absolute maximum level of association in respondent's responses based on the order of answering, and its respective k should indicate the length of the response pattern being repeated. In the next step, respondents are sorted with respect to their absolute maximum autocorrelation coefficient. Finally, the highest-scoring respondents, the number of which can be set in advance by the researcher, are selected for a closer inspection. During the closer inspection, the researcher must assess the validity of data for each respondent individually and use their own discretion on how to treat the data (most likely making a decision whether to keep, exclude, or partially exclude respondent's data). Needless to say, researcher's familiarity with both the data and the questionnaire they originate from is indispensable in this process.

Autocorrelation screening should be more sensitive to a repetitive response pattern the more it is at least partially repeated across respondent's answers. Any response pattern repeated multiple times yields perfect positive autocorrelation coefficient (r = 1) when k is equal to the length of the pattern, provided the pattern itself is uninterrupted over the whole string of responses. This works because the values are shifted by exactly one sequence length, making the repeated patterns match each other. In other words, each original value is correlated with the repeated identical copy of itself, resulting in a perfect fit.

When Autocorrelation is Not Available

Autocorrelation computation can sometimes fail and there are two reasons for this, both of them being associated with possible threat to data validity. The first reason is missing data. Unless the instrumental data matrix for a specific *k* provides at least two pairs of numeric values, autocorrelation is not available, as demonstrated in Figure 2, at the top. The second reason is associated with a presence of a long string of identical answers, the resulting zero variance for a specific *k* does not allow to calculate the autocorrelation, as illustrated in Figure 2, at the bottom. We recommend to keep track of the number of failed autocorrelations for each respondent and utilize it for more efficient screening.

Making Sense of Autocorrelation Coefficients

In order to be able to properly apply autocorrelation screening and interpret its results, users should have at least basic understanding of what kind of results autocorrelations produce on different data patterns. For this purpose, we present Figure 3 with some example data patterns and the resulting autocorrelation coefficients for each respective *k*.

Autocorrelation Screening in Practice

Data Format Requirements

Target data should be in numeric format and, as we mentioned above, arranged by the order of answering for each respondent.

Choosing Maximum *lag*

Choosing a suitable maximum *lag* (*k*) value, i.e. the maximum number of positions for the data to be shifted in autocorrelation analysis, is very important for a reliable screening. Maximum *k* value translates into the maximum length of a sequence within a repetitive response pattern that can be efficiently detected. Too low maximum *k* value hinders autocorrelation screening ability to detect longer repetitive response patterns, thus potentially lowering the method's sensitivity (the ability to correctly detect careless respondents). On the other hand, maximum *k* value set too high generally lowers reliability, because it makes the instrumental data matrix smaller, and, by calculating more autocorrelation coefficients, allows for higher frequency of occasionally strong autocorrelations that would inflate respondent's final autocorrelation score (determined as the highest absolute autocorrelation coefficient found), thus lowering the method's specificity (the ability to correctly not detect attentive respondents).

Figure 2. Two examples of data for which specific autocorrelation is not available

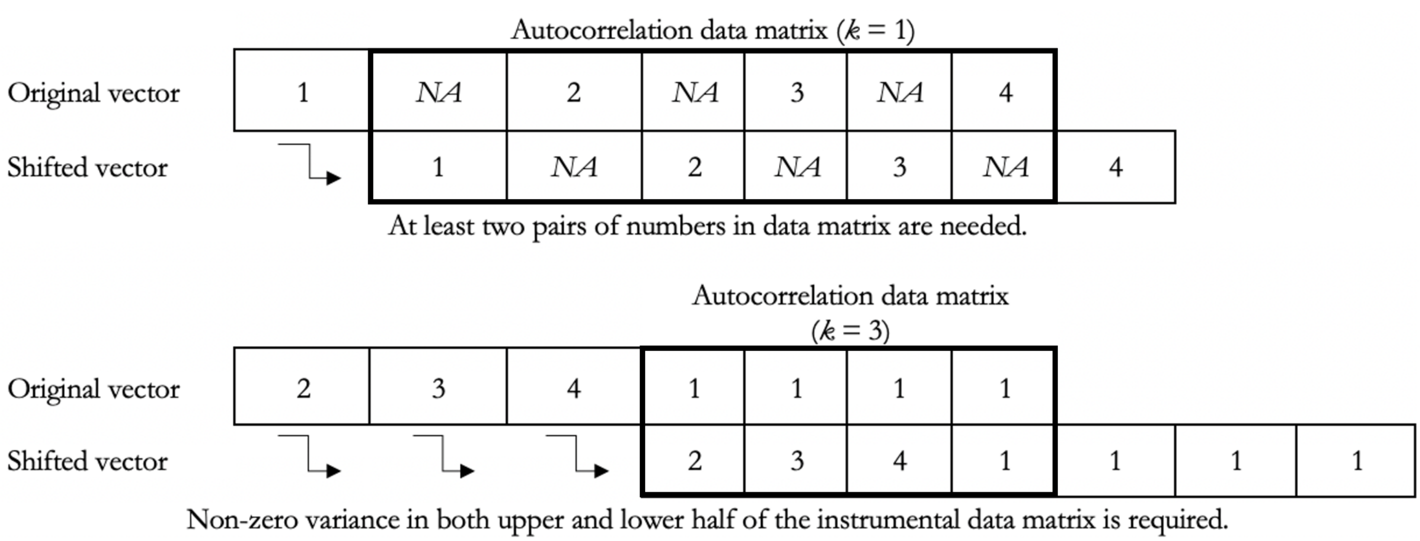
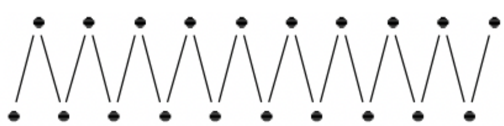


Figure 3. Autocorrelation coefficients for selected repetitive response patterns

Data pattern	*lag (k)* 1	2	3	4	5	6	7	8	9	10
1-2-3-4-5-1-2-3-4-5-1-2-3-4-5-1-2-3-4-5 	.12	-.45	-.54	-.11	**1.00**	.17	-.42	-.55	-.15	**1.00**
1-1-2-2-3-3-4-4-5-5-1-1-2-2-3-3-4-4-5-5 	.68	.29	-.02	-.35	-.46	-.56	-.41	-.25	.27	**1.00**
1-2-3-4-5-4-3-2-1-2-3-4-5-4-3-2-1-2-3-4 	.66	-.03	-.68	**-1.00**	-.65	.06	.70	**1.00**	.66	-.05
1-5-1-5-1-5-1-5-1-5-1-5-1-5-1-5-1-5-1-5 	**-1.00**	**1.00**	**-1.00**	**1.00**	**-1.00**	**1.00**	**-1.00**	**1.00**	**-1.00**	**1.00**

We recommend researchers to set maximum *k* value as the maximum length of the repetitive response sequence they expect to occur in their data. Maximum *k* value can be also modified post hoc after reviewing the autocorrelation screening results. Much fewer respondents with invalid data detected than expected may hint at improperly chosen maximum *k* value, warranting a new autocorrelation screening. Overall, we consider maximum *k* values within the range of 5–12 to be suitable for most questionnaire data, as a rule of thumb. We consider generating and correctly following a repetitive response pattern with the length of 12 or more to be quite demanding regarding respondent's attention, which goes against the fundamental concept of careless responding. Such long patterns should therefore be extremely unlikely to occur during careless responding.

Choosing Cut-off Screening Criteria

In order to efficiently screen for careless respondents, a researcher must set criteria for selecting potentially invalid respondents. These criteria should be set in accordance with the three main indicators: 1. the magnitude of the highest absolute autocorrelation coefficient, 2. the number failed autocorrelations per respondent, and 3. the amount of data the researcher is capable to closely inspect afterwards.

Highest Absolute Autocorrelation Coefficient. Unfortunately, it is not possible to determine what should constitute a "normal" value and what should be considered too high or potentially suspicious. The reason is that autocorrelation heavily depends on the design of the questionnaire itself. For example, questionnaires with highly correlated items naturally result in higher autocorrelations in respondents' data, compared to questionnaires with lower inter-item correlations. Similarly, the questionnaire factor structure, item order, and the presence of reversed items might increase or decrease autocorrelations, depending on circumstances. Therefore, an autocorrelation coefficient of, e.g. .60 cannot be considered "too high" or "too low" unless we compare it with the autocorrelation coefficient of other respondents on the same questionnaire. As a result, we recommend not to rely on absolute, but rather on relative criteria when assessing autocorrelation magnitude. In practice, that means researchers should select a certain percentage of the top-scoring respondents with regards to their autocorrelation coefficient. Inspecting a histogram of the highest absolute autocorrelation coefficients for all respondents beforehand might be tremendously helpful for a decision where to set the cut-off value.

Theoretically, some respondents with the lowest autocorrelation coefficients might be also worth inspecting in detail, especially if their scores are much lower than for the rest of the sample, suggesting an extremely deviant response pattern compared to the other respondents. Practically, this would mean that the respondent in question provided answers that are mutually much more independent than other participants' answers. Autocorrelation screening should be capable of detecting these cases as well and users are encouraged to always consult the histogram of the highest absolute autocorrelation scores in order to assess whether there are irregularities in the distribution of the autocorrelation scores either towards the high or low end of the axis. However, we must remind that autocorrelation score only serves to highlight potential validity problems and it does not justify data exclusion on its own. A deletion of either a high-scoring or a low-scoring respondent must be always done only after close data inspection and reasonable justification.

Autocorrelations Not Available. As explained earlier (see When Autocorrelation is Not Available), relatively high number of failed autocorrelations means that the respondent has either many missing data or they provided long strings of identical answers. Both cases spell a potential threat to validity of researcher's data and results. We recommend the same procedure as with the autocorrelation coefficients – sorting respondents by the number of failed autocorrelations and inspecting a percentage of the top-scoring ones. Nonetheless, cleaning the data beforehand with respect to the rate of missing answers is advised for two reasons: (a) relatively high missing data rate might on its own indicate a problem with answer validity, and (b) autocorrelation screening might often fail with circa 80% or higher missing rate, which is usually already too high for the respondent's data to be used anyway. To summarize, autocorrelation screening is not sensitive to missing data and it tends to fail only for very high non-response rates or long strings of identical responses. Consequently, autocorrelation screening is usable, albeit crude, tool for detecting careless respondents based on their non-response rate. For such purposes, we recommend using a dedicated missing data analysis. For the purpose of detecting highly homogeneous responding, we recommend analyzing response variance or using long string indices.

Amount of Data for Closer Inspection. Respondents screened out based on their autocorrelation coefficient must be individually inspected before any data are excluded, but close inspection of data is a time-demanding task. Therefore, researchers should consider how large portion of the most suspicious data they are able to manually check and evaluate with regard to their validity. This applies especially to large datasets of $N > 10{,}000$, where even 1% ($n > 100$) of the respondents could be too many to check one-by-one. Therefore, cut-off should be set also with regard to the final number of screened out cases and the amount of time a researcher can allocate to the data inspection task.

Practical Tools for Autocorrelation Screening

As a practical outcome of this paper, we have built a Shiny web application with graphical interface that allows for easy data upload, quick autocorrelation screening analysis, and comprehensive results overview. The said application also contains a brief tutorial on performing autocorrelation screening and is freely available to use online at https://jargottfried.shinyapps.io/Autocorrelation_screening/ or to download at https://osf.io/2h6m8/.

Additionally, we published a standalone R package *responsePatterns* (Řiháček & Gottfried, 2021) dedicated to autocorrelation screening. Compared to the Shiny application, this package allows for easier data handling, more adjustments, and offers a few additional results, as well as an alternative method of iteratively searching for repeating patterns of answers. We recommend the Shiny application to users inexperienced in R software, while for those with at least basic understanding of R software, the cited R package might be the preferred choice.

Important Factors in Screening Efficiency

Repetitive Response Pattern Length. Autocorrelation screening requires the pattern to be repeated at least once. As a result, shorter patterns (e.g. 1-3-5) can be repeated more times than longer ones (e.g. 1-1-2-2-3-3-4-4-5-5) and so the shorter repetitive response patterns can be detected more reliably. Notably, because autocorrelation screening essentially measures how much response variance can be explained by previous responses, this method is expected to poorly detect those who provided highly homogeneous response patterns with very low variance (e.g. 1-1-1-1-1-2-1-1-1-1). For the detection of these cases, we recommend using other existing methods like variance analysis, long string analysis, or multivariate outlier analysis.

Repetitive Response Pattern Interruption. Autocorrelations, just as standard correlations, can handle missing data quite well, provided that the underlying response pattern is not interrupted. That means that missing data should not shift the pattern. E.g. an original pattern of 1-2-3-4-5 being repeated with occasional missing data such as 1-*NA*-3-*NA*-5 should be robust even against relatively high missing data rates, but the repetitive pattern could be compromised if repeated as 1-*NA*-2-*NA*-3, instead.

Overall, in order to be able to detect repetitive response patterns, these patterns should not be disrupted and should not greatly change during answering (e.g. an interruption occurs when a respondent stops following the repetitive pattern for a few question, then resumes it). Disrupted or changed repetitive response patterns are much harder to detect through autocorrelation screening, with the method's impaired sensitivity being proportional to the extent of the repetitive pattern disruption. As such, autocorrelation screening can be expected to reliably identify the greatly careless respondents who stick to one repetitive pattern for the most of the questionnaire, but is likely to perform poorly in detection of respondents who answered some parts of the questionnaire carefully and some parts carelessly. This implies that autocorrelation screening would be suitable for detecting only careless respondents with a rigid response strategy, because respondents who change their repetitive responding pattern or respond carelessly only occasionally are unlikely to produce inflated autocorrelation coefficients. This fact greatly limits the practical usefulness of autocorrelation screening, because it has the potential to be useful in detecting only a portion of careless respondents. This method is not meant to be a substitute for a full-fledged inspection of data validity. Nonetheless, it should serve as a quick and easy-to-use addition.

Baseline Carelessness of Respondents. Because autocorrelation screening is most sensitive to heavily careless respondents, its application might be most useful and efficient on data gathered from easily distracted respondents like children, but the screening might fail to detect any suspicious respondents in attentive and highly motivated samples. Arguably, the context of the questionnaire administration could play a minor role as well, with much fewer heavily careless respondents being present in data gathered via face-to-face administration than via online administration.

Questionnaire Length. Since autocorrelation screening method relies on the repetitive response pattern to be repeated and reliably detects only those respondents who more or less consistently follow their pattern, questionnaire length plays a major role. Too few questions in a questionnaire means longer patterns cannot be sufficiently repeated and having too many questions raises the probability of the occasional pattern disruption. Generally, we recommend analyzing only 10–40 questions at one time. Analyzing fewer or more questions would probably lead to greatly inhibited sensitivity, i.e. the lower ability of the method to successfully detect careless respondents. In case the questionnaire is much longer than 40 questions, researchers should either choose a number of adjacent questions in which they expect the carelessness to be the most easily detected (like questions near the end of the questionnaire, because of respondents' possible fatigue), or they should split their data and conduct multiple autocorrelation screenings, each with appropriate number of questions (e.g. split a total of 80 questions into two datasets: 1^{st}–40^{th} question and 41^{st}–80^{th} question).

Scale Format. In order to prevent bias, only questions with the same answer scales should be analyzed at one time, ideally. Analyzing answers on two scales with vastly different number ranges together (e.g. answers on scale 1–5 and answers on scale 1–100) can bias the results to a great extent. Naturally, questions with unique scales or answer options where repetitive response patterns are unlikely or even impossible to emerge, like questions about gender or education, should be excluded prior to screening.

Factor Structure. Questionnaire factor structure determines the overall magnitude of autocorrelation coefficients. One-dimensional factor structure or multidimensional structure with strongly correlated factors can produce much stronger autocorrelations for respondents in general. Question order with respect to the factor structure can also greatly enhance or inhibit autocorrelations across all respondents. However, autocorrelation screening should overcome most of these effects, because it focuses on autocorrelation magnitude relative to the magnitudes for the rest of the sample. Nevertheless, we advise researchers to always consider questionnaire factor structure when evaluating validity of answers, since specific questionnaires may promote emergence of seemingly suspicious response patterns which can be actually perfectly valid with regard to the question content. To illustrate, one-dimensional questionnaire with answer scale 1–5, every odd question being reverse-coded, and all questions having approximately the same difficulty might produce answer patterns like 1-5-1-5-1-5-… for respondents very low or high in the measured trait. These particular respondents provide what we would consider to be a repetitive response pattern and they would also attain high autocorrelation coefficients. But in this case the answer repetitiveness does not imply low data quality, because the very structure of the questionnaire makes the repetitiveness theoretically plausible.

Overview of Method's Strengths and Limits

Compared to existing methods for inspecting data quality, we perceive the strengths of autocorrelation screening in (a) simplicity and fastness to compute, (b) no requirement of dedicated items or recording of response time, and (c) high sensitivity to repetitive response patterns.

On the other hand, autocorrelation screening is limited mainly by (a) narrow scope – it is sensitive only to certain manifestations of careless responding, (b) rather poor theoretical knowledge available about repetitive response patterns as a form of careless responding, and (c) lack of empirical evidence about its performance on genuine questionnaire data.

Conclusion

Autocorrelation screening has a potential to be efficient at detecting careless respondents who provide repetitive response patterns. The main advantage of this method is that it is quick and relatively simple to perform and interpret. Moreover, we provide a Shiny web application at https://jargottfried.shinyapps.io/Autocorrelation_screening/ alongside with a downloadable R package *responsePatterns* at https://CRAN.R-project.org/package=responsePatterns to allow researchers and broader public to easily perform autocorrelation screening on their own data. We argue this method could be useful for enhancing data quality by identifying certain careless respondents in psychological and sociological questionnaire research. However, due to the method having been just recently developed, we cannot yet empirically prove its sensitivity and specificity as a screening test for detecting careless respondents. This method was developed and tested using simulated data only.

In this paper, we laid down the theoretical rationale and methodical foundations to allow for the method's empirical evaluation. We propose that future research should focus on estimating the performance of autocorrelation screening based on inter-rater agreement on multiple questionnaire datasets, as well as specifying criteria under which the method performs optimally.

References

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *51*(4), 819–852. https://doi.org/10.1080/713755788

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Krosnick, J., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, *1996*(70), 29–44. https://doi.org/10.1002/ev.1033

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Řiháček, T., & Gottfried, J. (2021). *responsePatterns: Screening of careless responding patterns* [R package version 0.1.0]. https://CRAN.R-project.org/package=responsePatterns

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge University Press.

Towse, J. N., & Cheshire, A. (2007). Random number generation and working memory. *European Journal of Cognitive Psychology*, *19*(3), 374-394. https://doi.org/10.1080/09541440600764570

Towse, J. N., & Valentine, J. D. (1997). Random generation of numbers: A search for underlying processes. *European Journal of Cognitive Psychology*, *9*(4), 381–400. https://doi.org/10.1080/713752566

Treisman, M., & Faulkner, A. (1987). Generation of random sequences by human subjects: Cognitive operations or psychological process?. *Journal of Experimental Psychology: General*, *116*(4), 337-355. https://doi.org/10.1037/0096-3445.116.4.337

Zabelina, D. L., Robinson, M. D., Council, J. R., & Bresin, K. (2012). Patterning and nonpatterning in creative cognition: Insights from performance in a random number generation task. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 137–145. https://doi.org/10.1037/a0025452

Citation:

Gottfried, J., Ježek, S., Králová, M., & Řiháček, T. (2022). Autocorrelation screening: A potentially efficient method for detecting repetitive response patterns in questionnaire data. *Practical Assessment, Research & Evaluation*, 27(2). Available online: https://scholarworks.umass.edu/pare/vol27/iss1/2/

Corresponding Author:

Jaroslav Gottfried
Department of Psychology
Faculty of Social Studies
Masaryk University
Czech Republic

Email: jaroslav.gottfried [at] mail.muni.cz