Outrageous observations: The redheaded stepchild of data analysis

*What are outliers.*

Throughout psychology’s assessments of experimental procedures, participant observations have intrigued, inspired, and confused experimenters world wide. Some of these observations exist in such extremes that we begin to think of them as outliers, influential observations, or fringliers (Osborne & Overbay, 2004; Wainer, 1976). Since they were first commented on in 1777 by Bernoulli, outliers have been defined as many things. Munoz-Garcia, Moreno-Rebollo, Pascual-Acosta (1990) have synthesized their own definition that we feel is concurrent with our work:

“An outlier is an observation which being atypical and/or erroneous deviates decidedly

from the general behavior of experimental data with respect to the criteria which is to be

analyzed on it.” (pg 217)

Outliers can be observed in many forms, which should be outlined to ensure proper understanding and handling of such observations. Researchers have separated outliers into categories in many different ways over the years (Beckman & Cook, 1983; Hodge & Austin, 2004; Munoz-Garcia, Moreno-Rebollo, Pascual-Acosta, 1990; Orr, Sackett, & Dubois, 1991; Osborne & Overbay, 2004). Many of the categories used to describe outliers have noticeable overlap. For instance, Munoz-Garcia, Moreno-Rebollo, and Pascual-Acosta (1990) stated that outliers come about by errors in the way we gather data (inappropriate techniques, or experimenter error), errors in preparation (improper hypothesis, planning, or methods), or natural variability. Shortly after, Orr, Sackett, and Dubois (1991) described outliers as consisting of people being included in an experiment who aren’t part of the population, legitimate datapoints that are interesting because they do not fit the expected scheme, extreme datapoints on error components, human error in observation/coding, and errors in data preparation. Similarly, Hodge and Austin (2004) state that outliers arise, “due to mechanical faults, changes in system behaviour, fraudulent behaviour, human error, instrument error or simply through natural deviations in populations.” Osborne and Overbay (2004) have also specified that outliers follow from data errors, intentional or motivated mis-reporting, sampling error, standardization failure, bad assumptions about distributions, or legitimate data points. Taken together, we have delineated the following 4 types of outliers for the purpose of this paper.

1. Data entry error, experimenter error in processing, coding, or preparing data, as well as motivated misreporting.
2. Participants who do not represent the intended population. For example, a participant who uses a DVORAK or ergonomic keyboard will not be appropriate in a study investigating the typing speeds of typical QWERTY keyboard users.
3. Valid data points which bring to light interesting phenomena the study was not aimed at capturing. For instance, the 1854 cholera epidemic in London was found by Dr. John Snow to stem from a specific water well on Broad Street. While the majority of individuals infected lived very close to the contaminated well, Dr. Snow also found an instance of cholera far away from the well. When he rode out to investigate this outlying case, he found that the woman in question had someone ride to that specific well to draw water for her because she liked the taste (Vinten-Johansen, Brody, Paneth, Rachman, & Rip, 2003).
4. Natural deviations in the population.

*Effects of outliers*

These outliers can have serious effects on data, which can lead to imprecise data analysis (Osborne, & Overbay, 2004), confusing results (Stevens, 1984; Yuan & Bentler, 2001), and inappropriate conclusions (Coin, 2008; Stevens, 1984). By keeping outliers in a dataset, analyses are more likely to have increased error variance (depending on sample size, Orr et al., 1991) and biased estimates (Osborne, & Overbay, 2004) as well as reduced effect size and power (Orr, Sackett, & Dubois, 1991; Osborne, & Overbay 2004) which can alter the results of the analysis and lead to falsely supporting (Type I error) or denying a claim (Type II error). Additionally, incorrect estimates of effect lead to misleading meta-analyses or sample size estimates for study planning. Beyond these effects on analyses and conclusions these outliers can be inspirational to researchers and to their research models as they can encourage the diagnosis, change, and evolution of a research model (Beckman & Cook, 1983). All together, these issues caused by outliers can lead to furthering unwarranted avenues of research, ignoring important information, and creating erroneous theories, all of which serve to weaken the sciences.

*Determination of outliers*

*Reporting of outliers*

For these reasons, we aim to stress the importance of properly identifying outliers, their potential cause(s), and ways to handle the extreme responses.

Further, as outlined by the APA (American Psychological Association, 2010), one should acknowledge these steps not only at the time of analysis, but also at the time of publication. As Kruskal (1960) stated, no matter how we treat outlying observations,

“It is very important to say something about such observations in any but the most summary report. At least how many observations were excluded from the formal analysis, and why, should be given.” (pg. 1)

However, outlier report rates in studies have indicated that researchers either fail to acknowledge outliers or fail to report their acknowledgment of outliers (Huffcutt & Arthur, 1995 ; Orr, Sackett, & Dubois 1991) Huffcutt and Arthur (1995) looked at Organizational Behavior/Human Resource Management meta-analytic literature since 1987. As they were dealing with meta-analyses, the outliers they were identifying consisted of study coefficients, or correlations, not individual data points, as outliers are usually thought of. In their study they found only six of fifty meta-analyses used any outlier treatment for study coefficients. Also, in most of these, highly subjective means were used to determine which study coefficients were outliers. Similarly, Orr, Sackett, and Dubois (1991) inspected 100 Industrial/Organizational Psychology personnel studies and found no mention of outliers.

Although the APA and many researchers (Orr, Sackett, & Dubois, 1991, …) have instructed researchers to outline their data cleaning, there has been a dearth in recent literature investigating if these instructions are actually being followed. Therefore, the authors felt that the current study examined the results in 27 journals and 672 articles across 13 psychological disciplines to assess if outliers are being reported. Given previous report rates (Huffcutt & Arthur, 1995; Orr, Sackett, & Dubois, 1991) we expect to find little to no reporting of outliers in these articles. Further, we will examine how outlier reports relate to field, journal, analysis, and sample size. We will also investigate the methods and rational of outlier detection and removal.

**Method**

*Fields*

A list of psychological field areas was created to begin the search for appropriate journals to include. The authors brainstormed the list of topics (shown in Table XX) by first listing major research areas in psychology (i.e. cognitive, clinical, social). Second, a list of common courses offered at large universities was consulted to add to the list of fields. Lastly, the American Psychological Association’s list of divisions was examined for any potential missed fields. The topic list was created to capture large fields of psychology with small overlaps (i.e. cognition and neuropsychology) while avoiding specific subfields of topics (i.e. cognition, perception, and memory).

*Journals*

Once these topic areas were decided, researchers used various search sources (Google, EBSCO host databases) to find journals that were dedicated to each broad topic. Journals were included if they appeared to publish a wide range of articles within the selected fields. A list of journals, publishers, and impact factors (as noted by the journal website) was created for each field. Two journals from each field were selected based on the following criteria: 1) high impact factors, 2) impact factors over one a minimum, 3) a mix of publishers if possible, and 4) availability due to university resources. These journals are shown in Table XX.

*Articles*

Fifty articles from each journal were examined for data analysis. Data collection of articles started at the last volume publication from 2012 and progressed backwards until fifty articles had been found. We excluded online first publications and started in 2012 to ensure time for errata and retraction of articles. Articles were including if they met the following criteria: 1) included data analyses, 2) included multiple participants or data-points, and 3) analyses were based on human subjects or stimuli. Therefore, we excluded theory articles, animal populations, and single subject designs. Based on article review, three fields were excluded. Applied behavior analysis articles predominantly included single-subject designs, evolutionary psychology articles were primarily theory articles, and statistics related journal articles were based on user created data with specific set characteristics. Since none of these themes fit into our analysis of understanding data screening with human subject samples, we excluded those two fields from analyses.

*Data Processing*

Each article was then reviewed for key components of data analyses. Each experiment in an article was treated separately. For each experiment, the type of analysis, number of participants/stimuli, and if they indicated outliers were coded. Types of analyses were broadly coded into basic statistics (descriptive statistics, z-scores, t-tests, and correlations), ANOVAs, regressions, chi-squares, nonparametric statistics, modeling, and Bayesian/other analyses. First, we coded if outliers were mentioned at all in article. If so, we coded outliers into four types: 1) people, 2) data, 3) both, and 4) none found. We found that a separate coding for data was important for analyses with response time studies where individual participants were not omitted but rather specific data trials were eliminated. Then, the author decision on what to do with the outliers was coded into whether they removed participants/stimuli or left these outliers in the analysis, as well as if they tested the analyses with and without these outliers for determination of their effect on the study. If they removed outliers, a new sample size was recorded. Lastly, we coded the reasoning for outlier detection as either participant based (i.e. 3-month old infants were too fussy to be included), experimenter based (i.e. the experimental session was interrupted), statistical based (i.e. three SD from the mean), and no listed reason.

Table XX

*List of Fields, Journals, and Impact Factors*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field | Journal | Publisher | Impact Factor 2012 | Impact Factor 2017 | Proportion Change (d value) | Number of Experiments |
| Applied Behavior Analysis | Journal of Experimental Analysis of Behavior | Wiley | 1.39 |  |  |  |
| Applied Behavior Analysis | Journal of Applied Behavior Analysis | Wiley | 1.19 |  |  |  |
| Clinical | Journal of Consulting and Clinical Psychology | APA | 4.85 |  |  |  |
| Clinical | Journal of Clinical Psychology | Wiley | 2.12 |  |  |  |
| Cognitive | Cognitive Psychology | Elsevier | 4.27 |  |  |  |
| Cognitive | Journal of Experimental Psychology: Learning, Memory, and Cognition | APA | 2.85 |  |  |  |
| Counseling | Journal of Counseling | APA | 3.23 |  |  |  |
| Counseling | The Counseling Psychologist | Sage | 1.82 |  |  |  |
| Developmental | Journal of Experimental Child Psychology | Elsevier | 3.12 |  |  |  |
| Developmental | Journal of Youth and Adolescence | Springer | 2.72 |  |  |  |
| Educational | Journal of Educational Psychology | APA | 3.08 |  |  |  |
| Educational | Contemporary Educational Psychology | Elsevier | 2.20 |  |  |  |
| Environmental | Journal of Environmental Psychology | Elsevier | 2.93 |  |  |  |
| Environmental | Environment and Behavior | Sage | 1.27 |  |  |  |
| Forensics | Psychology, Public Policy, and Law | APA | 1.93 |  |  |  |
| Forensics | Law and Human Behavior | Spring | 2.16 |  |  |  |
| Industrial Organizational | Organizational Behavior and Human Decision Process | Elsevier | 3.94 |  |  |  |
| Industrial Organizational | Personnel Psychology | Wiley | 2.93 |  |  |  |
| Neurological/Physiological | Neuropsychology | APA | 3.82 |  |  |  |
| Neurological/Physiological | Cognitive, Affective, and Behavioral Neuroscience | Springer | 3.57 |  |  |  |
| Social | Journal of Personality and Social Psychology | APA | 5.08 |  |  |  |
| Social | Journal of Experimental Social Psychology | Elsevier | 2.31 |  |  |  |
| Sports | Journal of Sport & Exercise Psychology | Human Kinetics | 2.66 |  |  |  |
| Sports | Sociology of Sport Journal | Human Kinetics | 1.00 |  |  |  |
| Overview | Psychonomic Bulletin & Review | Springer | 2.25 |  |  |  |
| Overview | Psychonomic Science | Sage | 4.43 |  |  |  |
| Overview | Psychological Assessment | APA | 2.99 |  |  |  |

*Note.* Impact factors as of time of data collection (Spring 2013).

* Analyses:
  + Do we mention outliers at all? (yes/no)
  + Does this reporting of outliers have any relationship to the specific discipline?
  + Journal?
  + Type of analysis?
  + Sample size?
  + When we do talk about them, what rational do we use to take them out most often?
  + Do we remove people or data points?
  + Do we test with and without them to determine if they should be taken out?
* Discussion: Importance of published material addressing outliers:
  + Helps us become a cumulative science
  + Helps us understand differences in effect sizes
  + Ties into the replication project ideas.

Vinten-Johansen, P., Brody, H., Paneth, N., Rachman, S., & Rip, M. (2003). *Cholera, chloroform, and the science of medicine: A life of John Snow.* New York : Oxford University Press.