Practical decision-making in psychology research

Math 1530: Introduction to Statistics

Overview

The purpose of a psychology research study is to find and report those effects within a sample randomly drawn from a population, that are likely to reflect what is present in the actual, larger population. Because it is rare to collect data from every person in a population, instead, in most cases, psychologists use statistics to determine how confident they can be that differences observed in the sample truly exist in the population. Using this process, statistics support or refute inferences that could be drawn about the *parameters* (numerical facts) of the population.

When humans try to explain their observations of others, they jump to biases and wrong conclusions much of the time. For this reason, it is important in research to use the scientific method and statistical theory to study behavior. Specifically, using randomized procedures and testing observations with statistical analyses before making inferences about a population from a sample utilizes the established mathematics of chance to draw more objective conclusions.

However, in practice, when researchers design studies and then later when they use statistics to analyze the data they collected, they often confront decisions that are not completely straightforward. This is true, even when the researchers do their best to conduct an objective study. Three practical types of non-straightforward scenarios that are often encountered by psychologists will be discussed in this document: sample-related biases, missing data, and publication bias. Each of these aspects can impact on the results from a research study, and if that happens, they can impact on the conclusions drawn about the whole population. But techniques and ways of approaching these issues are always under development in the field, as well as resources for open science practice.

Sample-related biases

In order for a study to reliably tell society something about a population, research participants (people who volunteer to participate in the study) must be selected in a representative manner from the population that is being studied[a] and also assigned to experimental conditions equally. In a controlled experiment, it is possible to assign research participants roughly equally to different experimental conditions, as long as the researcher can use a random process to do so, like a random number generator, an alternating protocol or a coin toss. It is more difficult to find and recruit participants in a fully representative way. Three common reasons for this are described in hypothetical (plausible, but fictional) examples below.

One reason is that participants often tend to opt to participate based on reasons that aren't related to the research question. In the first example, the goal is to learn something about three-year-old children living in Boston. A hypothetical study was conducted at Boston University. Parents were recruited through preschools and city programs for families. The parents volunteered to bring their child to the university during a weekday to participate in the experiment. Later, the researcher found that nearly all of the families who came to the university to participate reported middle class incomes. This might have happened because families in which at least one parent typically had flexible weekday time and an extra vehicle were a lot more likely to bring their child to the university on a weekday. Because the requirements for participating meant that children were not sampled at random, the results of the study were not representative of most three-year-olds living in the Boston area.[1]

The second reason is that it's often easier to find and recruit research participants at certain places where some people are a lot more likely to be present than others. In the second example, a hypothetical study was run in a lab at a college in the US. Psychology students were given extra credit for participating in the study. At the end, there were 50 participants. The mean age was 19, and the age range was 18-23. So, the youngest participants were 18 years old and the oldest participants were 23 years old. The results of this study do not reflect adults at different ages and the study didn't include any adults who did not go to college. So, the sampling for the study was not representative if the goal was to understand adults living in the United States[2-3].

Finally, studies themselves are often more accessible to some people than to others[4]. In the third example, a hypothetical study on losing parents was conducted through social media, and a link was posted to online support groups for bereavement. Participants clicked on the link to participate, where they indicated some information about themselves, including their age and gender, when their parent had died, and then typed answers to interview-style questions via an online form. But the form was presented and the responses had to be made in English without another language option. In addition, only people who participated in support groups for bereavement were likely to find out about the study. In this case, there are again two reasons why the sampling was not representative. First, it was easier for fluent English-speakers to participate, and second, people who had lost their parents but who were not part of any online bereavement support group were unlikely to find out about the study.

---

[a] In theory, research participants should be selected from a population at random, but due to location and other logistical constraints, researchers might aim for a local sample that is reasonably representative of the population they aim to study, sometimes applying randomized or clustered selection within one or more specific geographical regions. The important factor is that deciding who to sample is carried out in an impersonal manner.

But in reality, the reasons for sample-related biases often extend beyond *sample selection*. Immediately below are summaries of two well-known cases of sampling problems in past research studies that show this. The first example summarizes clinical trials in the US in 1954 in which polio vaccines were given to children. In this example, we will see why a randomized controlled trial is more controlled than assignment to a treatment or control group based on year in school. The second was of an observational study versus a randomized controlled trial of ultrasounds during pregnancy and birthweight in the late 1980s. In this example, we will see why observational studies are not experiments.

Historical example 1: Polio vaccination trials in the 1950s.

The information in this section and the results presented in Table 1 were summarized directly from an example given in Chapter 1 of a well-known statistics handbook.[5]

The first widely reported epidemic of polio occurred in the US in 1916. By the 1950s, more than one vaccine had been invented, but none had been proven effective. In 1954 the vaccine developed by Jonas Salk was tested across the US. Because rates of polio varied from year to year, the only way to test the effectiveness of the vaccine was to give the vaccine to some children, while leaving others unvaccinated.

Parents had to provide permission (consent) for their child to be vaccinated as part of the research. So, the study could have been designed so that the children whose parents consented were assigned to the treatment group, while the children whose parents did not consent were assigned to the control group. But it was known that parents with larger incomes were more likely to consent to the vaccine than were parents with smaller incomes. The design would be biased against the vaccine, because the children in larger-income households were more vulnerable to polio. Income was connected to housing conditions, and a child who lived in less hygienic surroundings was more likely to contract a mild case of polio in early childhood, while still protected by antibodies from their mother, and develop stronger long term immunity as a result. So, to be able to establish how effective the vaccine was, it was necessary to assign children to the treatment or control group, not based on whether their parents consented to their vaccination, but by an impersonal factor.

Table 1

The results of the Salk vaccine trials of 1954[5b]. The table shows the size of the groups and the rates of polio per 100,000 in each group, rounded.

Randomized	controlled	study	School year as group study		
Group	Size	Rate	Group	Size	Rate
Treatment	200,000	28	Grade 2 (vaccine)	225,000	25
Control	200,000	71	Grades 1 & 3 (control)	725,000	54
No consent	300,000	46	Grade 2 (no consent)	125,000	44

Note. The original source for the table was Francis (1955). An evaluation of the 1954 poliomyelitis vaccine trials-summary report. *American Journal of Public Health, 45*, 1-63[5].

As the Table 1, shows, in the end, two different types of designs were conducted to compare vaccinated with unvaccinated children-depending on location. Neither approach used the “no consent” group as the control group. The table shows a summary of the rates of polio contracted by children who were assigned to different groups in the randomized controlled study (left, colored in blue) and in grade-school comparisons (right, colored in peach).

[b] The HeLa cell line was used in researching a vaccine for polio[6].

Only children in the treatment group were vaccinated, while the children in the control group were not given the vaccine. In the school-year-as-group design (Table 1, right), the vaccine and control groups were different sizes, but the *rate* of polio (the number of cases per 100,000 children) was compared between the groups. This ratio (summarized as the Rate columns in Table 1) adjusted for the different sizes of the groups.

In the randomized controlled experiment (left, colored in blue), children were randomly assigned to either a treatment or a control group. Randomly means that a coin could be tossed for each child, and based on whether it landed heads or tails, the child would be assigned to one of the groups. The children in the treatment group received a vaccination for polio, while the children in the control group received an inactive injection (a placebo). Neither the children, their families, nor the staff that diagnosed polio, knew which injection each child had received.

The study on the right (colored in peach) was less well controlled. In this study, the control group was not like the treatment or like the no-consent group, because the control group, grades 1 and 3, contained both families that would have consented and that would not have consented to the vaccination.

The mathematics of chance processes guarantee that if enough participants are assigned to each experimental condition randomly, the resulting groups will closely resemble each other in both known and unknown factors. In this case, because only the randomized controlled trial on the left compared children in the treatment and control groups who were as alike as possible, the observed difference between these groups gave a clear idea of the strength of the vaccine.

Historical example 2: Do ultrasound exams cause low birthweight?

The information in this section was summarized directly from an example given in Chapter 2 of a well-known statistics handbook.[5]

During pregnancy, prenatal health is examined using ultrasound technology to image the uterus of the parent using low-power sounds waves. Yet, early experiments on animals in laboratories had found that ultrasound examinations caused low birthweight. An observational study was carried out at Johns Hopkins hospital in Baltimore around 1988 in order to find out whether ultrasounds caused low birthweight in humans. At the time, parents who had an ultrasound exam had pregnancies that differed from those who did not. Specifically, an ultrasound was recommended when a pregnancy had potentially complicating factors that could result in low birthweight. The researchers found an association between ultrasounds and low birthweight, even after controlling for all additional variables that they could. Around the same time, a large randomized controlled trial conducted from the UK showed that an ultrasound screening did *not* cause low birthweight[7]. Ultrasounds are now considered a safe screening procedure and are recommended during pregnancy because they provide important medical information. Observational studies are not experiments, and even when additional variables are controlled for as thoroughly as possible in an observational study, it is still harder to be certain about the results.

Summary

This section showed how sample-related biases can come up when finding participants for a study, when assigning participants to different experimental conditions (such as a treatment or control condition), and when trying to match samples in an observational study.

Randomized controlled designs are an effective solution to the problem of how to assign different people equally into different experimental conditions, but no solutions have been invented for sampling problems during sample selection, or for confounding variables in observational studies, that are guaranteed to work for nearly any type of study. Instead, there are partial solutions developed for particular types of studies (like choosing districts and households at random for polling the opinions of eligible voters or controlling for age during the analysis for an observational medical study) that help mitigate likely biases.

Further Reading

[1]Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31-38. https://doi.org/10.1016/j.jecp.2017.04.017

[2]Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, *28*(3), 450-461. https://doi.org/10.1086/323732

[3]Druckman, J. N., & Kam, C. D. (2011). Students as experimental participants. *Cambridge Handbook of Experimental Political Science*, *1*, 41-57. https://www.ipr.northwestern.edu/documents/working-papers/2009/IPR-WP-09-05.pdf

[4]Boas, T. C., Christenson, D. P., & Glick, D. M. (2020). Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods*, *8*(2), 232-250. https://doi.org/10.1017/psrm.2018.28

[5]Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (4th ed.). W.W. Norton & Company.

[6]Skloot, R. (2010). *The Immortal Life of Henrietta Lacks*. Crown Publishers.

[7]Waldenström, U., Nilsson, S., Fall, O., Axelsson, O., Eklund, G., Lindeberg, S., & Sjödin, Y. (1988). Effects of routine one-stage ultrasound screening in pregnancy: a randomised controlled trial. *The Lancet*, *332*(8611), 585-588. https://doi.org/10.1016/S0140-6736(88)90636-8

Missing data

In psychology research, most datasets end up missing some measurements. This can occur because some participants didn't complete the entire study, because there was an error and a measurement had to be omitted, or because of legitimately occurring outlying values that could not be otherwise explained by error. Often, finalized datasets are missing some data, in the form of empty spaces/cells in the columns of a table where the corresponding measurements would have been, for multiple different reasons[8-9]. Hopefully, the overall proportion of data that are missing will be small.

But missing data create problems for statistical analysis. First, it is not possible to compute the relationship between paired variables if one of the values from a pair is missing. For example, this can be a problem in a paired samples t-test or a repeated measures analysis of variance (ANOVA). Second, parametric statistical models typically require a normal distribution. Third, outlying values can change the results of statistical analyses. Researchers have to decide what to do with data that are already missing, when it is appropriate to omit data and how to adjust their analyses accordingly. Data should only be omitted from a study for external and objective, and never for personal and subjective, reasons.

Incomplete entries

Some observations never get recorded in the first place: for example, because a participant left questions blank on a questionnaire, because an experiment was stopped before it was completed, or because a measurement didn't get saved properly. In the Methods section of a published research paper, missing entries are typically reported as counts or as percentages and classed according to the reason for their absence. Researchers often choose to omit participants who do not have complete, or near-to-complete, entries from the study entirely. When participants are omitted from the study, this information is typically summarized in the Participants section[9].

Poor data quality

Another reason that measurements are omitted is because of poor data quality, even though this decision is made during or after the fact-that is, after at least some relevant data were collected. For example, an experimenter might have made a mistake while administering a test that makes the responses recorded doubtful, a participant might have responded by pressing the wrong button consistently throughout an online experiment, or a participant could have moved around too much during an MRI brain scan taken as part of a research study, making the recording of doubtful quality. In cases like these, the researcher will typically choose to omit doubtful measurements from the study. This is usually done based on widely used norms and, where possible, established rules, for what makes a measurement acceptable after possible corrections have been made, applied to each type of measurement used in the study. Often, this decision occurs after the data have been collected and before analyzing the data.

Outlying values

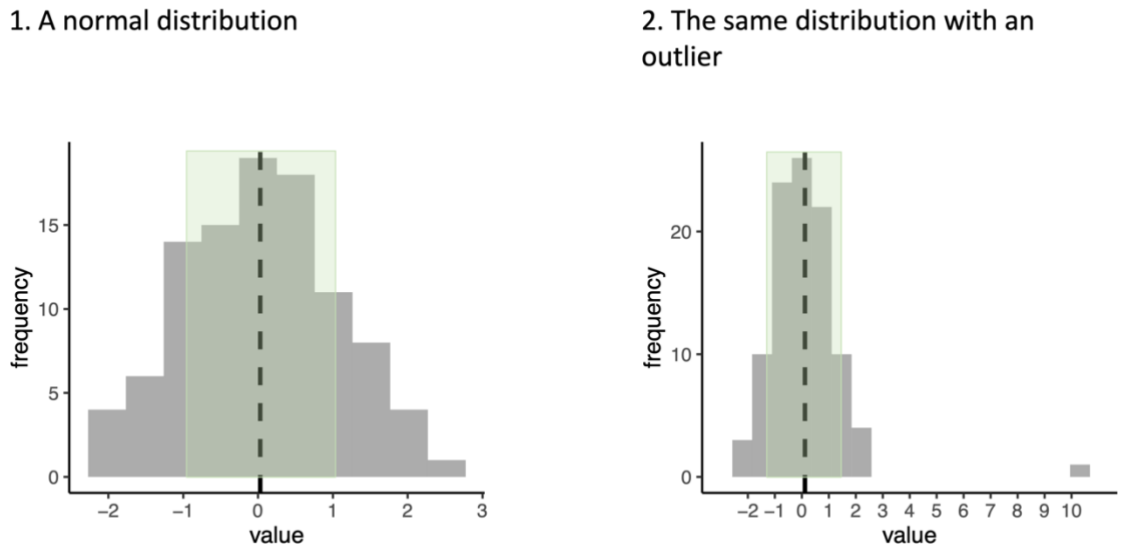
Even when data were recorded and saved and the data quality are acceptable, some measurements produce unusual values compared to the rest. Statistical outliers are values that fall at the extremes of a distribution of values and are often defined as values that fall more than +/- 2.5 standard deviations (SD)s from the sample mean, or more than +/- 1.5 times the length of the interquartile range (IQR) away from its closest border[8]. Chance

variability alone produces some outlying values during repeated measurements, proving that outlying values are just something that happens in research (Chapter 6[5]).

Statistical outliers can affect overall measures of central tendency and spread in a distribution, specifically the mean and standard deviation, so that the mean is pulled in the direction of the extreme value and the other values are more clustered together. An illustration of a normal distribution with and without an outlying value is shown in figure 1. For this reason, boxplots that show the median and interquartile range instead of the mean and standard deviation are usually used to identify which values are outliers, because in the case of boxplots, the presence of an outlier does not dramatically change the median and interquartile range (figure 2). As is illustrated theoretically in figure 3, outliers can also bias statistical results through impacting the slope of the regression line, that shows the average change in a unit of y with a unit of change in x.

Figure 1

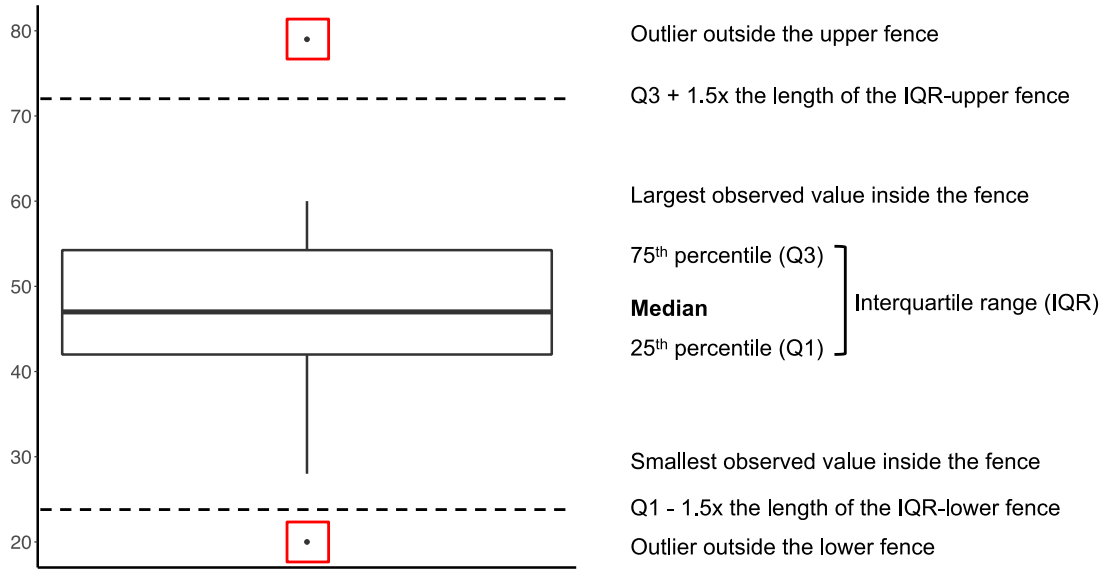
Illustration of an outlying value within a histogram



Note. A randomly created normal distribution with 100 values is shown on the left (1), and the same distribution with an outlier is shown on the right (2). In both figures, the area within one standard deviation of the mean is highlighted in green. A small change can be seen between the means of the distributions, from .03 on the left to .11 on the right. With a reasonably large number like 100 values, a single outlying value is unlikely to drastically change the mean. The standard deviation is also more spread out, encompassing values between -1.02 and 1.02 in the normal distribution on the left, and -1.42 and 1.42 on the distribution with an outlier on the right.

Figure 2

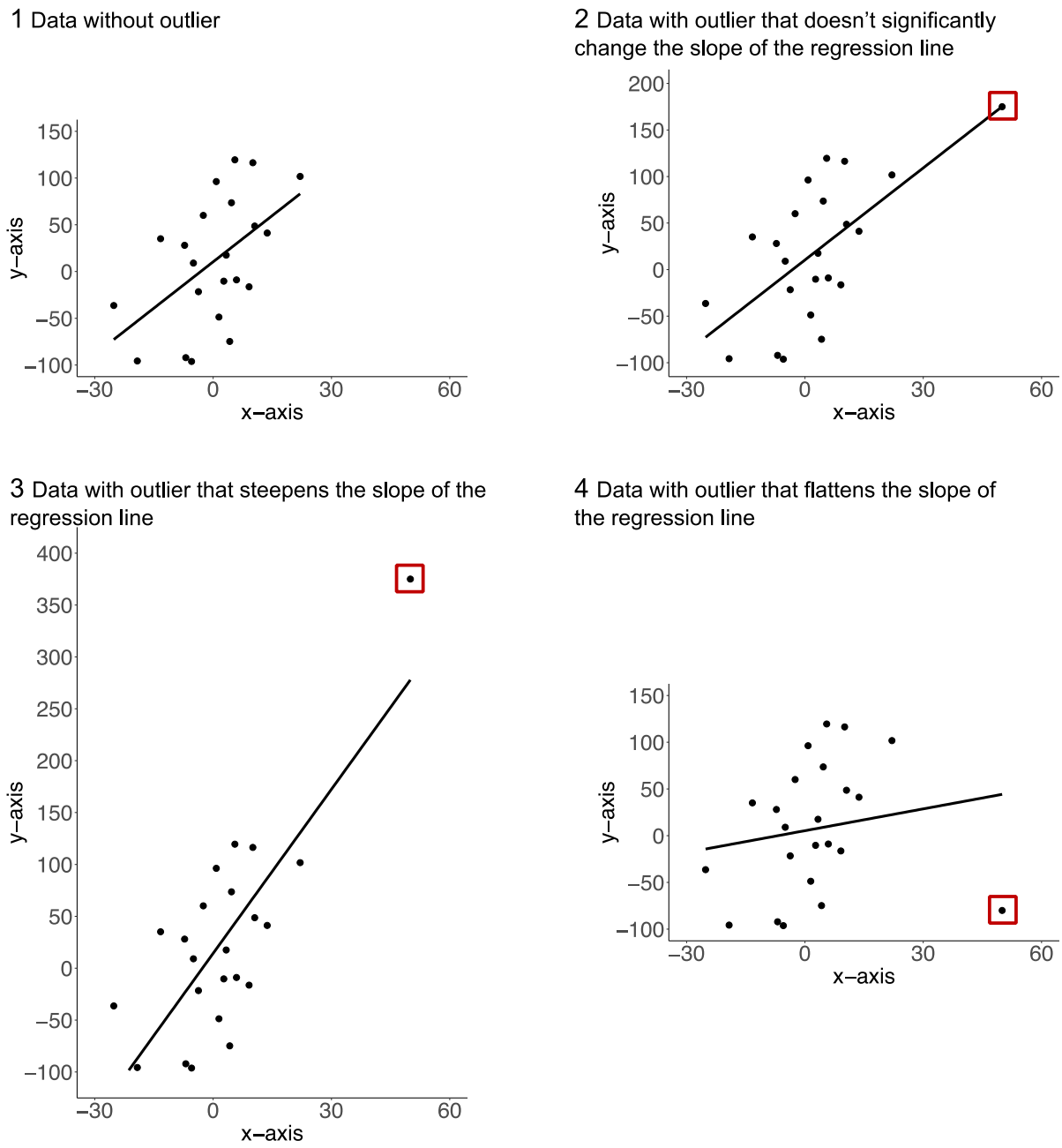
Illustrations of outlying values within a boxplot



Note. Boxplot statistics are usually used to identify outlying values. In this example, outlying values are outside of the upper and lower and fences, shown as dotted lines. The interquartile range (Q3-Q1) is 12 points on the number line. The upper fence is 54 (Q3) + 1.5*12 = 72. The largest observed value is 79, an outlier outside of the fence. The lower fence is 42 (Q1) - 1.5*12 = 24. The smallest observed value is 20, outside of the fence. The parts of the boxplot are labelled on the right.

Figure 3

Examples of how outlying values impact a regression line



Note. Three examples show how different outlying values at x = 50 2-4) may theoretically relate to the same pattern of remaining data 1), either remaining consistent with 2) or changing the slope of 3-4) a regression line. Changes in the ranges of the y values between the figures show the differences. The remaining data were randomly generated with r = .6.

The outlying values in 2-4) have high *leverage*, where the value in the independent variable (x) falls far away from the other x-values. However, not all of them change the slope of the regression line in a linear model. The outlying value that does not significantly change the slope of the regression line 2) does not change the theoretical relationship between the x variable (the independent variable) and the y variable (the dependent variable). But this

comparison has some limitations, because little is known about the correct theoretical relationship between x and y at such a large value of x. If the correct theoretical relationship turns out to be nonlinear, then the outlying value in 2) is *not* representative, and instead, the shape of the model should be changed.

In the case of observations where there isn't a concern that an outlying value is different from the others because of experimental error, some researchers choose to leave these values in the dataset when conducting a statistical analysis, because no matter why the values are omitted, the reason to omit them will be arbitrary on some level. Other researchers choose to omit outlying values when using a linear regression model, because in most cases, the remaining values will give a better general representation of the underlying population. (However, this becomes less and less of a concern with larger samples; to read more about the *central limit theorem*, see Chapter 18 in[5]). Still other researchers keep outlying values that do not significantly influence the slope or the fit of the regression (closest-fitting) line but omit values that do (see figure 3)[10]. In practice, this issue, and the opinions that individual researchers hold about it, is more complicated than can be fully described here[10-11].

There isn't a universally agreed-upon answer as to whether to omit statistically outlying values. Some questions that researchers consider when making this decision are: do these values reflect something important to understand about the population (a reason to keep them), are they likely to bias conclusions drawn from the sample about the population (a reason to omit them), and do they cause the data to violate any mathematical assumptions for the statistical test used (a reason to omit them)? But when in doubt, the safest way is to report results from both versions of a statistical analysis: a version with and without outliers.

Further reading

[8]Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology*, *70*(4), 407-411. https://doi.org/10.4097/kjae.2017.70.4.407

[9]Praharaj, S.K. & Ameen, S. (2021). Writing the methods section in a manuscript. (Column: Tips on Research and Publication). *Kerala Journal of Psychiatry, 34*(1), 79-83. https://kjponline.com/index.php/kjp/article/view/277/304

[10]Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270-301. https://doi.org/10.1177/10944281124708

[11]Grace-Martin, K. To drop or not to drop. *The Analysis Factor*. https://www.theanalysisfactor.com/outliers-to-drop-or-not-to-drop/

Publication bias and open science practice

In order for the scientific method to operate as intended, the research process must remain *transparent* and *reproducible*. Transparent means that the complete research process is available to another investigator. Reproducible means that when a study is repeated by different investigators, similar results are observed. Publication bias can impact both transparency and reproducibility. An important example of publication bias is the preference of research journals to publish studies that turned out in a way that seems "successful" to the editor, and to reject studies that turned out in a way that seems "unsuccessful." Specifically, research journals are more likely to publish work that includes an expected, statistically significant finding and to reject work that does not. Most researchers report *p* values to establish statistical significance, and therefore, to get published, a study must have found at least one *p* value < (less than) .05 that supports one of the researcher's *hypotheses* (predictions about what would be observed based on prior evidence and scientific theory). Publication bias has multiple negative consequences for psychological research.

First, when a researcher plans a new study, they may have limited means of finding out whether similar studies were attempted previously that did not find a significant difference in the comparison they plan to test. This is because journals were very unlikely to publish such work, and instead, viewed it as a failure. In the past, researchers had few ways to communicate about or share studies that did not turn out as expected with other researchers.

Second, publication bias places significant pressure on researchers to change their study plan after the fact in such a way as to get a $p < .05$ that supports one of their hypotheses. For example, researchers may change their hypothesis and/or the way they analyze their data, and then, out of several tries, report the approach that produced a statistically significant result without disclosing the exploration that took place up to that point. The test that is reported is accurate. If another person ran the exact same statistical test on the same data, they would get the same result. However, the final report does not include a record of the other attempted analyses that produced non-significant results. If another person read related studies and formed their own hypothesis about the same study and picked a statistical test to analyze the same data, they would be a lot less likely to obtain a statistically significant result. In short, because of publication bias, researchers experience pressure to discretely modify, or tweak, their study in order to obtain a statistically significant result that supports a hypothesis, because otherwise, the study is unlikely to be published.

Because of publication bias, proportionally more statistically significant results are published than are actually found in psychology studies, and it is less likely that published effects will reliably be found again in repeated experiments. All of this reduces confidence in the scientific process. To combat this, many more resources are now available as part of *open science practice* to enable psychologists to communicate about all of their experimental results, to specify their study plan ahead of time and to report the "story" of their study, including all of the changes made, not just the final results. Some of these resources are included below. This doesn't solve everything about publication bias, but presents a significant start.

Further Reading

A Wikipedia article on publication bias:

https://en.wikipedia.org/wiki/Publication_bias

Resources for open science

The Open Science Framework allows researchers to pre-register (pre-specify) their study plans, describe the changes they made to their study along the way, and share their study materials and data with other scientists:

https://osf.io/

https://www.youtube.com/watch?v=iebMBpi0prc

Registered reports are a relatively recent practice to be adopted by many research journals. The journal accepts a researcher's study plan, and agrees to publish the study ahead of time, no matter how the results turn out:

https://www.cos.io/initiatives/registered-reports