16 Statistics 3

In the previous two discussions of statistical tools for psychological science, we reviewed the main workhorses of experimental research: the t-test and the ANOVA. Most designs that have one or more variables manipulated across two (or more) levels with a measured operational definition of the dependent variable will be analyzed with those tools. Here we will review two additional simple statistical methods that can be applied to relationships between experimental variables when it is not the case that we have a categorical independent variable and a measured dependent variable.

The correlation analysis is a mathematical tool for quantifying the reliability of the relationship between two continuous variables. It is very commonly used in non-experimental designs, although less so in experimental research. It does serves as an introduction point to a set of more complex approaches that are used in non-experimental research. At the end of the chapter we will very briefly review a set of more complex statistical tools that build on linear regression models with the observation that correlation is the simplest possible version of linear regression. Detailed discussion of the use and application of these tools are beyond the scope of this text but are core to complex quantitative techniques used to help draw conclusions about causality between variables in non-experimental design.

We will also describe a non-parametric statistical tool that is applied to cases where the dependent variable is a categorical outcome rather than a continuous measure. The **chi-squared** χ^2 analysis assesses the reliability of the effect of manipulating an independent variable on the rates of occurrence of a dependent variable with two distinct outcomes. The name of this analysis references the statistical parameter used to characterize the strength of the effect of the IV on the DV, just as in our other statistical tools, except that in this case the statistical parameter is the Greek letter χ, pronounced ki. This analysis method is presented as an introduction to a class of analysis tools for this kind of outcome data. As with general linear regression models, we will not provide a thorough exploration of these more complex tools here.

Learning Objectives

1. Understand when and how to use a correlation analysis
2. Interpret and understand ranges for Pearson's and Spearman's r
3. Understand how correlational data is displayed in a scatter plot diagram
4. Understand when and how to do a X2 analysis with categorical data
5. Understand data presented in rates tables
6. Understand how to draw inferences from X2 analysis.

Correlation Analysis

For non-experimental research, simple percentages may be computed to describe the percentage of people who engaged in some behavior or held some belief. But more commonly, non-experimental research involves computing the correlation between two variables. A **correlation coefficient** describes the strength and direction of the relationship between two variables. The values of a correlation coefficient can range from −1.00 (the strongest possible negative relationship) to +1.00 (the strongest possible positive relationship). A value of 0 means there is no relationship between the two variables. Positive correlation coefficients indicate that as the values of one variable increase, so do the values of the other variable. A good example of a positive correlation is the correlation between height and weight, because as height increases weight also tends to increase. Negative correlation coefficients indicate that as the value of one variable increase, the values of the other variable decrease. An example of a negative correlation is the correlation between stressful life events and happiness; because as stress increases, happiness is likely to decrease.

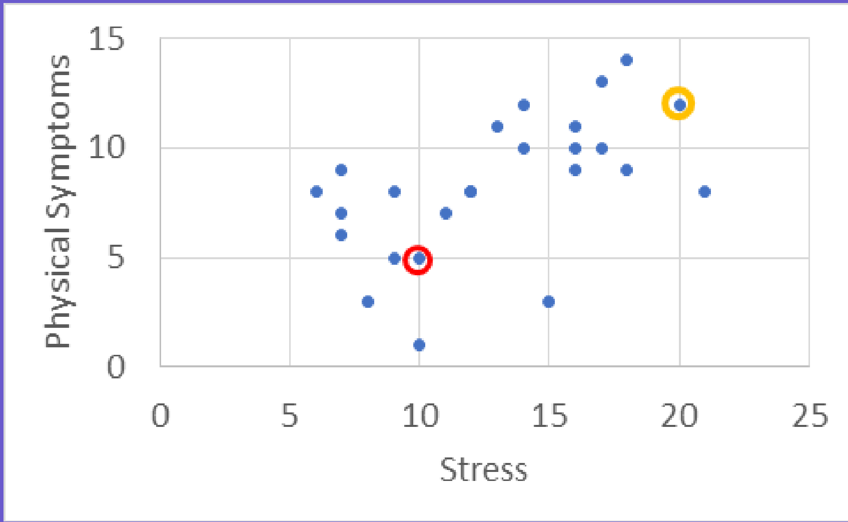
The phrase ***Correlation is not Causality*** is common and important to psychological science. The challenges this creates for validity of conclusions was reviewed in detail in Chapter 14. Here we consider the statistical tools used to assess reliability of the relationship between two variables. As always, a reliable relationship does not automatically imply the conclusion is valid. We might observe a very robust relationship between variable X and Y and still not know for sure if X caused Y, Y caused X or some third variable Z caused the observed relationship between X and Y.

Mathematically, a correlation analysis simply assesses the strength of the relationship between two continuous variables. In theory, if the independent variable was manipulated across a continuous range by the experimenter, we could draw perfectly reasonable causal inference from a correlation coefficient. In practice, this is vanishingly rare, so it is generally safe to use the heuristic that if you see a correlation coefficient reflecting a correlation analysis, the research to which it is being applied is likely also correlational research and non-experimental.

Correlations Between Quantitative Variables

Correlations between quantitative variables are often presented using **scatterplots**. An example is shown on the right based on hypothetical data on the relationship between the amount of stress people are under and the number of physical symptoms they have. Each point in the scatterplot represents one person's score on both variables. For example, the circled point in red represents a person whose stress score was 10 and who had five physical symptoms. The orange circled point is a participant with a stress score of 20 and twelve physical symptoms. Taking all the points into account, one can see that people under more stress tend to have more physical symptoms. This is a good example of a positive relationship, in which higher scores on one variable tend to be associated with higher scores on the other. In other words, they move in the same direction, either both up or both down. A negative relationship is one in which higher scores on one variable tend to be associated with lower scores on the other. In other words, they move in opposite directions. There is a negative relationship between stress and immune system functioning, for example, because higher stress is associated with lower immune system functioning.

Scatter Plots

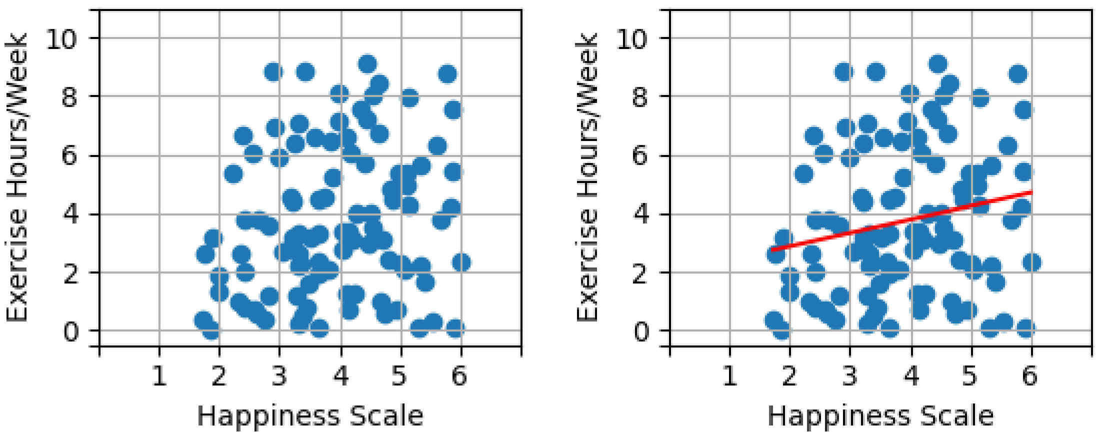


The graph type shown to the right is a *scatterplot*. This is a graph where each data point is shown on (x,y) coordinates. This type of data visualization is notable for making visible every point in the dataset, which can be messy but is very useful for spotting outlier points and getting a sense of the overall relationship between the two measures. Many tools also allow a quick addition of a *trendline* to help visualize the linear relationship.

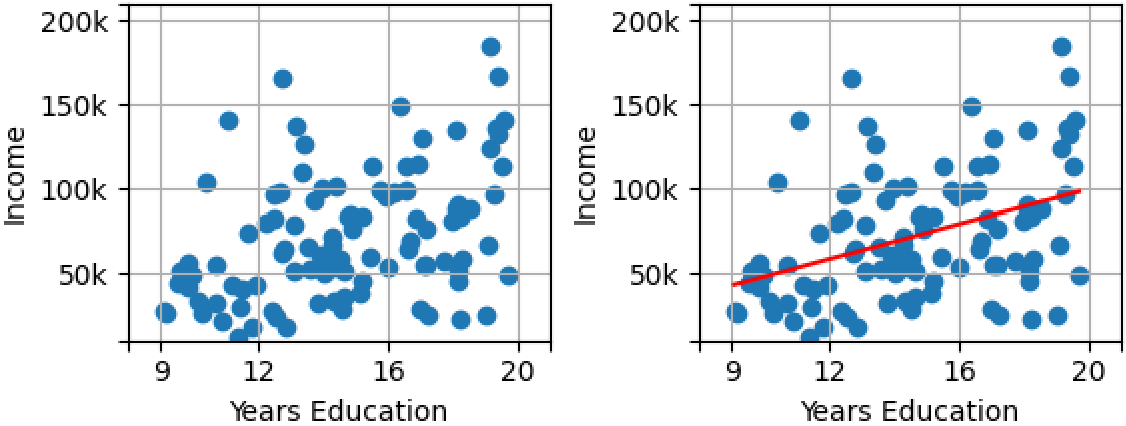
The strength of a correlation between quantitative variables is typically measured using a statistic called **Pearson's Correlation Coefficient** (or **Pearson's r**). Pearson's r ranges from −1.00 (the strongest possible negative relationship) to +1.00 (the strongest possible positive relationship). A value of 0 means there is no relationship between the two variables. When Pearson's r is 0, the points on a scatterplot form a shapeless "cloud." As its value moves toward −1.00 or +1.00, the points come closer and closer to falling on a single straight line. Correlation coefficients near ±.10 are considered small, values near ± .30 are considered medium, and values near ±.50 are considered large. Notice that the sign of Pearson's r is unrelated to its strength. Pearson's r values of +.30 and −.30, for example, are equally strong; it is just that one represents a moderate positive relationship and the other a moderate negative relationship. With the exception of reliability coefficients, most correlations that we find in psychology are small or moderate in size. The scatterplot above has a correlation coefficient between the hypothetical data of 0.55, a fairly strong positive relationship where physical symptoms (y-axis) go up as stress (x-axis) go up.

An analysis that produces a correlation coefficient is expressed with the statistical parameter, ***r***, which like other statistical parameters (t, F) reflects the strength of the relationship between the variables. It is also associated with a p-value, which always has the same definition, the probability of observing this relationship by chance if the null hypothesis was correct. For a correlation analysis, the null hypothesis is that there is no relationship between variables which would produce an r = 0.00.

In this simulated example, we see a relationship between hours per week of exercise and scores on a happiness questionnaire. The correlation shown here is r = 0.20. The left panel shows the scatterplot of all the data (100 simulated values). With such a weak relationship, it is difficult to see that happiness is going up very slightly with increased exercise. On the right panel, a trendline has been added which shows the average increase. For this kind of analysis, it should be noted that the weakness of this correlated relationship is affected by the number of points that are very far from the trendline. Although there is a tendency for points to be higher on the right side of the graph, the relationship is not very strong, although it does meet the reliability criterion of p<.05.

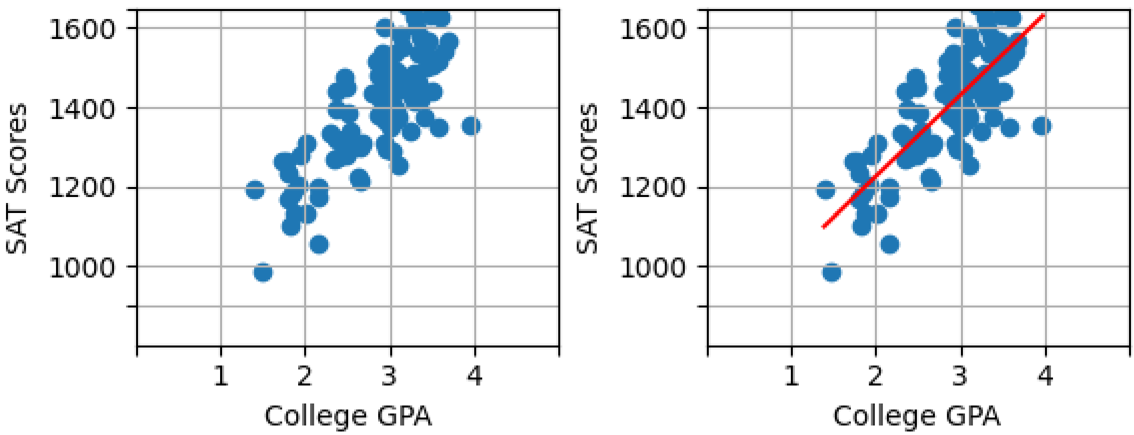


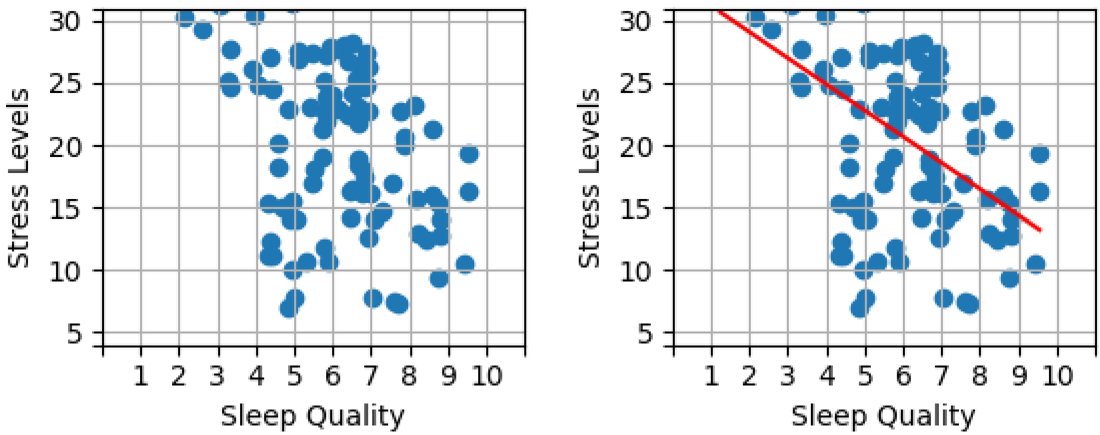
r
=



In the second example above, we have a simulated relationship between years of education and income after school. On the left panel, the relationship is now more easily visible, reflecting the fact that this relationship is producing a correlation of r = 0.4. The relationship is even easier to see on the right panel with the added trendline. You might note that the slopes of the line is not that different from the previous example, but the dots (which are each one participant) are more closely clustered around the line.

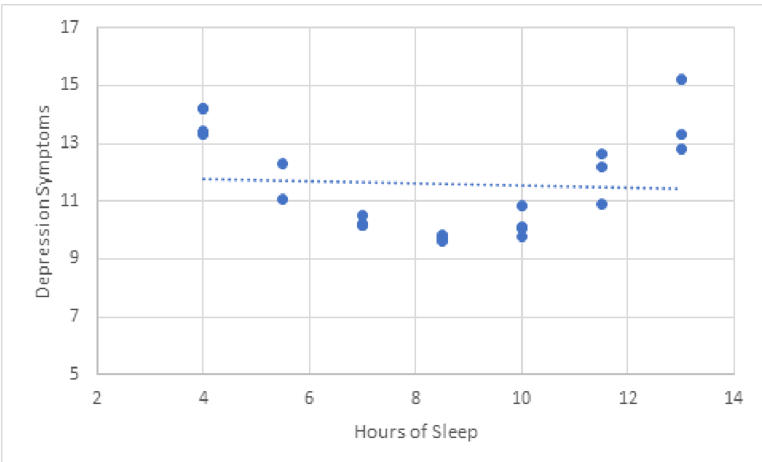
The example below shows a much stronger relationship between college grades and standardized tests taken before starting at school. These simulated data show a very strong relationship, r = 0.80.





The previous three examples all showed positive relationships between the two variables assessed. In this example, we see a negative correlation reflecting that sleep quality decreases with higher stress levels. These simulated data illustrate a robust negative relationship, r = -0.5.

There are two common situations in which the value of Pearson's r can be misleading. Pearson's r is a good measure only for linear relationships, in which the points are best approximated by a straight line. It is not a good measure for curvilinear relationships, in which the points are better approximated by a curved line. The figure below, for example, shows a hypothetical relationship between the amount of sleep people get per night and their level of depression.



In this example, the line that best approximates the points is a U-shaped curve because people who get about eight hours of sleep tend to be the least depressed. Those who get too little sleep and those who get too much sleep tend to be more depressed. Even though the figure shows a fairly systematic relationship between depression and sleep, Pearson's r would be close to zero because the points in the scatterplot are not well fit by a single straight line (flat trend line shown). This means that it is important to make a scatterplot and confirm that a relationship is approximately linear before using Pearson's r. Curvilinear relationships are fairly common in psychology. The technique for measuring them extends the idea of the linear trendlines seen on the scatterplots above to curvilinear lines defined by polynomials, which goes beyond the scope of the tools discussed here.

Another common situations in which the value of Pearson's r can be misleading is when one or both of the variables have a limited range in the sample relative to the population. This problem is referred to as restriction of range. Assume, for example, that there is a strong negative correlation between people's age and their enjoyment of hip hop music as shown by the scatterplot across age ranges from 18 to 80. However, if data were collected from a restricted range sample, e.g., 18 to 24, the relationship might not be visible. This is yet another example of why we cannot confidently draw conclusions from null results. It is also a reminder that calculation of a correlation coefficient based on Pearson's r depends on having data sampled across a reasonably wide rand and also assumes that the distribution of both the x and y variables are roughly normal (following a Gaussian distribution).

A tool to be aware for conditions in which the data are not normally distributed is the **Spearman's rank correlation**. This also results in calculating an r statistic that acts just like the Pearson's correlation. Spearman's correlation can be used when the observed data has a number of notable outliers that would not be expected in a normally distributed dataset. This correlation coefficient is calculated based on ranking the data such that the lowest value is recoded as 1 and each higher value is one more so that the highest value in the data set is the number of total participants.

This reduces the distorting impact of extreme outliers that can reduce the effectiveness of a more typical Person's correlation. An example of where this tool can be used effectively is in the analysis of reaction time (RT) data where most of the responses cluster around some average speed but there are a few extremely slow responses (a very common shape of RT data). This produces a highly skewed distribution that is not Gaussian (normal). A rank correlation, Spearman's, analysis enables analysis of these types of data without problems caused by the violation of the assumption of normality. As with all tools that allow us to carry out statistical analysis when data violate assumptions of normality, this should be used with caution and some thought towards why the data are not normally distributed.

Analysis of Categorical Data

In our standard model of experimental design, we use a manipulated (experimental) independent variable and measure a dependent variable. The IV (or factors) typically have a small number of levels, often 2, among which participants are assigned. We can think of our IV as being defined by a categorical variable in that participants are assigned to one condition or the other. The DV is a continuous variable that we can then look for differences in the average score across conditions. The correlation analysis described above is one variation from this model where both the IV and DV are continuous measures. It can also be the case that the DV sometimes needs to be a categorical variable.

The canonical examples of categorical variables can be captures in the memorable phrase, *you can't be a little bit pregnant or a little bit dead*. These are events for which there are only two outcomes: you are, or you aren't. Measures of these kinds of variables are 'binary' in that there are two possibilities. It is also possible to have categorical variables for which there are more than two alternatives. In general, if the alternatives can be ordered in a systematic, ranked way, these will often be coded as a familiar continuous variable. But there are plenty of cases where there is a range of options that are each independent choices. For example, one might look at some aspect of high school education and what college within Northwestern a student applied with the possible outcomes being WCAS, McCormick, Medill, or the Bienen School of Music. Here the outcome variable is categorical across four possibilities.

These approaches can very well be experimental, and the same concepts drive our ability to draw inferences from the data: did the IV affect the DV? But now we need a statistical tool that allows us to characterize how the shift in categorical choices was affected by the manipulated IV. The general approach for analyzing these data is to organize the outcomes into a **contingency table**.

As an example, consider a non-experimental study that asked athletes if they generally stretch before exercising and if they have had an injury in the past year. Whether or not they stretch is a categorical variable with two possibilities: yes/no. The same is true for whether they have had an injury in the past year: yes or no. Suppose we had data from 800 athletes. We could organize the results in the following contingency table.

	Injury	No Injury	Total
Stretches	**55**	**295**	350
Does not stretch	**231**	**219**	450
Total	286	514	800

The bolded data in the table reflects the actual data counts and the rightmost and bottom columns are sums of the data above. Looking at the outcome counts, we can see that the number of people with injuries is much lower in the group that stretches (55 versus 231), but we should also note that there are different numbers of participants in the stretch/no-stretch conditions (350 versus 450). To correct for this, we should calculate **rates** for all the conditions here to see if there is evidence the injury rate is different for the two stretching conditions. The rate of injury for the stretches condition is 55/350 = 15.7%. The rate of injury the not stretching condition is 231/450

= 51.3%. That is clearly a lot higher, but what we have done so far is effectively calculated the descriptive statistics for a categorical design, which are usually rates at which the outcome occurred. To draw conclusions about whether stretching causes a lower rate of injuries, we need a statistical test to identify if this difference is statistically reliable. As with all other statistical approaches, we want the familiar p-value, which is always the probability of observing this pattern of data under the null hypothesis regardless of the kind of data being evaluated. The null hypothesis here is that stretching or not are associated with injuries at the roughly the same rate overall and the observed different was somehow just luck.

There are a variety of ways to analyze a contingency table, but we will focus on one tool that is fairly flexible across common experimental designs, the chi-squared analysis or X^2. The X is the Greek letter 'chi' (pronounced ki) and the analysis is variously referred to by the Greek letter, the term *chi-squared* or the mixed term *chi²*. Conceptually, this analysis is based on looking at the difference between the **observed rate** of occurrence from the **expected rate** of occurrence in each of the four cells of the contingency table. The expected rate is the rate of occurrence under the null hypothesis that stretching does not matter.

It is useful to note that the injury rate in the dataset does not have to be 50% just because there are two possibilities. We can estimate the average injury rate by looking at the total number of injuries over all the participants ignoring the stretching condition (just as we did with marginal means for main effects in ANOVA). There were a total of 286 injuries out of the 800 participants, which is a rate of 35.8%. The X^2 formula is essentially telling us that if the average rate of injury is 35.8%, what are the odds that one group would exhibit a 15.7% rate and the other group a 51.3% rate. The number of participants is critical for this calculation, so it is actually done with the numbers expected in each cell not from the rates themselves. The X^2 is the sum of the difference between the Expected number of participants in each cell (under the null) and the Observed number of participants. For the stretches condition, at a 35.8% injury rate, we should have seen ~125 injuries, yet we only saw 55. X^2 is actually calculated as the square of this difference divided by the expected value and so the sum of this across the four cells is:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

For this example, we would obtain a X^2 value of 108.7, which is exceedingly improbable and since $p<.05$ (the same criterion as always), we can reject the null hypothesis and conclude that stretching did reliably affect the injury rate in this data collection sample.

Nothing about the calculation would be different if this was an experimental study where athletes were assigned to a stretching or no stretching condition and then injury rate was assessed afterwards. Of course, that would be a problematically unethical study since we would have assigned participants to a condition that we thought might cause them to become injured.

The method applies to any design where the DV is a categorical variable and works well even if there are more than two possible outcomes. The general approach is to organize the data into a contingency table of counts of each of the combinations of IV level and DV outcomes. From this, calculate the rates to be able to describe the data as descriptive statistics. The calculation of a X^2 analysis is simple enough that you could do it with a calculator or spreadsheet, but you can also simply use standard analysis programs like R which also helpfully provide the direct estimate of the p-value.

Extending this methodology to multiple independent variables is somewhat tricky and requires digging into a set of analytical tools known as **non-parametric statistics**. These tools are useful for categorical data and situations where assumptions of normality are not met in the observed data but are not very frequently used in basic experimental science.

Advanced Statistics

Here we will briefly review a set of more complex analytic methods to provide some familiarity with the names of these approaches. We will not attempt to apply these methods in this class, but you may see references to these approaches in published research reports that you encounter as part of your background reading. These fairly common but more complex analyses include **regression**, **general linear models**, and **factor analysis**. We will also introduce the tool of meta-analysis, which aims to aggregate information from multiple published studies to help better quantify effect sizes and reliability across similar experimental approaches.

Relationships Among Multiple Variables

A common extension of research based on correlations between a single continuous independent variable and a continuous dependent variable is to collect data from a number of IVs and combine these to understand their relationship to the DV. The relationship between these IVs and the DV is assessed via a regression analysis, within a general framework of an approach called general linear modeling (GLM). The approach of using a GLM is the basis of the vast majority statistical modeling of complex datasets except for a handful of special and interesting cases where non-linear relationships are evaluated (although if you work in this area, you learn a variety of approaches for transforming non-linear data to be suitable for linear modeling).

The correlation analysis described above is a simplified case of regression where there is one predictor (IV) variable and one outcome (DV) variable. The trendlines shown in the correlation examples are calculated by a linear regression. Whereas in correlation, we get a statistical parameter r, GLM analysis will often report a statistic, r^2, which has a useful verbal description of **percentage of variance** accounted for in the data. This is, in fact, the same r statistic we use for correlation analysis squared and can help understand what a correlation analysis is telling us. The idea behind the phrase *percent of variance accounted for* is that the data have variance which results from a range of factors related to extraneous variables, participant variances and measurement error. Our predictor variable (IV) accounts for some of the variability in the measured variable (DV) but we acknowledge that there are many other sources of variance not accounted for. In the example of a fairly strong correlation of $r = 0.5$, $r^2 = 0.25$ meaning a quarter of the variance is accounted for and three quarters (75%) is still unknown. So even in the case of a strong relationship, there's still a lot we have not captured in the statistical model. For milder correlations in the range of 0.3, we would be happy to account for 10% of the variance. The r^2 statistic is generally reported when using any GLM (regression) analysis. Many ANOVA programs are using a GLM behind the scenes (ANOVA is also a special case of the broader regression approach) and may report the r^2 statistic as well. If so, you can use this heuristic to get a sense of how systematic your data are by how much of the variance is accounted for in your model.

Another important use of regression analysis is to explore possible causal relationships among variables even when these measures are collected in a correlational research style (without manipulating the IV). The primary way of doing this is through the statistical control of potential third variables. Instead of controlling these variables through random assignment or by holding them constant as in an experiment, the researcher instead measures them and includes them in the statistical analysis called **partial correlation**. Using this technique, researchers can examine the relationship between two variables, while statistically controlling for one or more potential third variables. You will typically see this described as *controlling for*, as in X appears to cause a change in Y controlling for Z in an analysis where the effect of Z are attempted to be controlled for to try to give an independent view of how X affects Y.

For example, assume a researcher was interested in the relationship between watching violent television shows and aggressive behavior but she was concerned that socioeconomic status (SES) might represent a third variable that is driving this relationship. In this case, she could conduct a study in which she measures the amount of violent television that participants watch in their everyday life, the number of acts of aggression that they have engaged in, and their SES. She could first examine the correlation between violent television viewing and aggression. Let's say she found a correlation of +.35, which would be considered a moderate sized positive correlation. Next, she could use partial correlation to reexamine this relationship after statistically controlling for SES. This technique would allow her to examine the relationship between the part of violent television viewing that is independent of SES and the part of aggressive behavior that is independent of SES. If she found that the partial correlation between violent television viewing and aggression while controlling for SES was +.34, that would suggest that the relationship between violent television viewing and aggression is largely independent of SES (i.e., SES is not a third variable driving this relationship). On the other hand, if she found that after statistically controlling for SES the correlation between violent television viewing and aggression dropped to +.03, then that would suggest that SES is indeed a third variable that is driving the relationship. If, however, she found that statistically controlling for SES reduced the magnitude of the correlation from +.35 to +.20, then this would suggest that SES accounts for some, but not all, of the relationship between television violence and aggression. It is important to note that while partial correlation provides an important tool for researchers to statistically control for third variables, researchers using this technique are still limited in their ability to arrive at causal conclusions because this technique does not take care of the directionality problem and there may be other third variables driving the relationship that the researcher did not consider and statistically control.

If you continue to study more complex research methods beyond this class, you will encounter regression analysis, ANCOVA (analysis of covariance), MANOVA (multiple variable ANOVA), logistic regression, and even more complex tools such as structural equation modeling. Many of these are used in non-experimental research studies to try to increase the confidence in drawing causal conclusions from datasets where the independent variables cannot be manipulated (e.g., epidemiology, economics). It may be useful to you then to know that these are all founded on the GLM core and developing an understanding of this multiple regression approach will help you grasp a wide range of more specialized tools used for those areas of research.

Multiple Dependent Variables

In the development of novel survey instruments, it is common to collect a set of dependent variables and look for relationships among these. This can be used to calculate both convergent and discriminant validity of the scale. For example, when Cacioppo & Petty (1982) first reported the Need for Cognition Scale—a measure of the extent to which people like to think and value thinking—they used it to measure the need for cognition for a large sample of college students, along with three other variables: intelligence, socially desirable responding (the tendency to give what one thinks is the "appropriate" response), and dogmatism. The results of this study are summarized in a correlation matrix (below) showing the correlation (Pearson's r) between every possible pair of variables in the study.

Correlation Matrix Showing Correlations Among the Need for Cognition and Three Other Variables Based on Research by Cacioppo and Petty (1982)

	Need for cognition	Intelligence	Social desirability	Dogmatism
Need for cognition	--			
Intelligence	+.39	--		
Social desirability	+.08	+.02	--	
Dogmatistm	-.27	-.23	+.03	--

For example, the correlation between the need for cognition and intelligence was +.39, the correlation between intelligence and socially desirable responding was +.02, and so on. (Only half the matrix is filled in because the other half would contain exactly the same information. Also, because the correlation between a variable and itself is always +1.00, these values are replaced with dashes throughout the matrix.) In this case, the overall pattern of correlations was consistent with the researchers' ideas about how scores on the need for cognition should be related to these other constructs.

A more complex statistical technique for evaluating relationships among a large number of variables is the approach of **factor analysis**. Factor analysis attempts to organize the observed data as arising from a smaller number of predictor variables than were originally used. As an example, researchers Rentfrow & Gosling (2008) asked more than 1,700 university students to rate how much they liked 14 different popular genres of music. They then submitted these 14 variables to a factor analysis, which identified four distinct underlying factors. The researchers called them Reflective and Complex (blues, jazz, classical, and folk), Intense and Rebellious (rock, alternative, and heavy metal), Upbeat and Conventional (country, soundtrack, religious, pop), and Energetic and Rhythmic (rap/hip-hop, soul/funk, and electronica). The underlying idea is that the rating of the blues, jazz, classical and folk music tended to be similar to each other, so these are reduced to one underlying factor. Note that this analysis does not tell you anything about what the factor means, which is often considered a weakness of this approach, leaving it up to the authors to decide to describe this cluster as *Reflected and Complex*. The analysis provides a table of *factor loadings* (right) which indicate how well each of the observed measures relates to the inferred cluster (factor).

The way to read a factor loadings table is to treat these as similar to a correlation. Participants who liked Blues tended to like other music that was scored as highly related to the Reflected and Complex factor: Jazz, Classical and Folk. This technique is often used to distill a large dataset with many measures into a description based on a smaller number of underlying factors.

Table of Factor Loadings of 14 Music Genres studied by survey of undergraduate students

Genre	Reflective and Complex	Intense and Rebellious	Upbeat and Conventional	Energetic and Rhythmic
Blues	.85	.01	-.09	.12
Jazz	.83	.04	.07	.15
Classical	.66	.14	.02	-.13
Folk	.64	.85	-.04	-.07
Rock	.17	.85	-.04	-.07
Alternative	.02	.80	.13	.04
Heavy Metal	.07	.75	-.11	.04
Country	-.06	.05	.72	-.03
Sound tracks	.01	.04	.70	.17
Religious	.23	-.21	.64	-.01
Pop	-.20	.06	.59	.45
Rap/hip-hop	-.19	-.12	.17	.79
Soul/funk	.39	-.11	.11	.69
Electronica/ dance	-.02	.15	-.01	.60

Survey data analyzed through Four Varimax-Rotated Principal Components. Based on Research by Rentfrow and Gosling (2003).

Meta-Analysis

The technique of **meta-analysis** describes a method for aggregating data across many published studies. As researchers have become more concerned about replicability and reliability of research, this tool has become increasingly popular to draw broad conclusions across many published studies rather than rely on individual reported experiments. The approach is typically carried out in a series of steps. First, the set of relevant published research studies is identified through a systematic literature search. Then for each publication, the reported effect size is identified, sometimes together with other methodological elements to quantify the quality of the study. Studies with larger samples or lower overall variance are likely to be reporting higher quality results. The effect sizes across many studies are then combined to identify the central tendency of the magnitude of the effect of the IV on the DV. This can simply be the mean of the effect sizes, but it can also be a weighted average to give more impact from the higher quality research. This approach provides an excellent way to get a good estimate of the true size of the effect of the IV on the DV.

Meta-analysis can also be used to look at the distribution of results, which can help identify outlier studies with findings that look very different from what should be similar reports. The distribution of effect sizes across studies can also be used to identify publication bias or whether the tendency across studies supports the null hypothesis that the IV does not affect the DV. These can be very helpful approaches in controversial research areas where findings do not consistently replicate. Comparison of methodology across findings can also help identify what extraneous variables might be affecting outcomes that were imperfectly controlled in individual studies.

The primary challenge of meta-analysis is that it requires having a lot of published research available to aggregate. As a result it can only be applied in research areas where several researchers have published related results. Good estimation of effect size would be very valuable for research in newer areas so that power analysis can be used to plan samples for research, but meta-analysis does not help here. The desire for more findings to incorporate in meta-analysis has led some researchers to call for relaxing the p-value criteria for reliability to allow studies with marginally reliable findings or even null findings to be incorporated into this approach.

Exercises

Question 1. A colleague helping with a research project comes to tell us in that the latest statistical analysis, they found an r = 2.30? Should we be excited?

Question 2. A college administrator is choosing between 2 admissions tests, one that correlates with subsequent college performance r = -.54, and one for which, r = +.45. Which test should be preferred and why?

Question 3. Of all the statistical tests we have covered in class so far, explain which is appropriate for each of the following hypothetical designs and why:

- Older people (age in years) are found to exercise less (in hours per week) than middle aged people
- At a car dealership, people planning to buy a car are either shown a documentary on climate change or not and then they are scored as to whether they purchased an electric vehicle or gas vehicle
- Hospitalization rate (percent of people hospitalized) due to COVID infection differ in counties that voted red (Republican) or blue (Democrat) in the 2020 election
- The relationship of first year salary post-graduation with average GPA in sophomore year during college
- Number of social media posts made by people who either took research methods in college or did not
- An analysis of scores on the Oxford Happiness Scale based on participants diet (vegetarian/not) and whether they have a Facebook account.

17 Qualitative Research

The history of psychology is often described as starting with William James, an American philosopher, who wrote **The Principles of Psychology** in 1890 that essentially established psychology as a field separate from natural philosophy. Although there was a developing tradition in experimental psychology being developed by German academics, especially Wilhelm Wundt, around the same time, James' work is notable for relying primarily on the technique of **introspection**.

Introspection largely relies on self-reflection and attempting to document the process of psychological experience by focus on subjective experience. While the overview of psychology that James produced using this method still contains a very effective overview of many aspects of the field, introspection as a tool became somewhat disfavored in the early part of the 20th century. It quickly became clear that much of the process of cognition that drives behavior is not particularly accessible to description based on subjective experience. Introspection also does not allow itself to rigorous study as there is no way to resolve differences in self-reflection across people.

From a methodological standpoint, introspection, observational and other qualitative approaches were replaced by **behaviorism**, championed by BF Skinner in the 1930s. The experimental tools used in this approach were early versions of the basic methodology described here but were often mainly applied to animal studies. Research with experimental animals avoids some of the challenges in developing operational definitions of complex human thought or behavior by keeping the dependent variables to simple and easily measured responses in animals.

The **cognitive revolution** in psychology in the 1960s can be seen as bringing human participants into experimental methodology. Research from that time and since has recognizable methodology of random assignment to different levels of an independent variable and subsequent measurement of a dependent variable. At around the same time, tools for more detailed physiological study of animal models led to the development of the field of **neuroscience**, which capitalized on a lot of the methodologies developed by behaviorists combined with sophisticated biological manipulations and measures. In Chapter 20, we will review some methods of **cognitive neuroscience** that attempts to bring those ideas back to the study of human participants again.

Within the framework presented here for psychological science, we can consider James' initial work with introspection to provide a set of terms for the constructs of psychological science. Introspection is too subjective for effective science, but we can approach these constructs by creating operational definitions to be used in rigorous scientific methods. Even in doing so, we should acknowledge that there are many areas of psychology in which we are still identifying the key constructs to be studied. In these areas, tools known as **qualitative research**, which rely more on observation, case studies, collecting stories, and structured interviews. In some cases, these data can be used to drive quantitative measures for research and in others, the observations themselves can illustrate directions in which experimental methods might be applied in the future.

Learning Objectives

1. List several ways in which qualitative research differs from quantitative research in psychology.
2. Describe the strengths and weaknesses of qualitative research in psychology compared with quantitative research.
3. Give examples of qualitative research in psychology.
4. List the various types of observational research methods and distinguish between each.
5. Describe the strengths and weakness of each observational research method.

What Is Qualitative Research?

This textbook is primarily about quantitative research, in part because most studies conducted in psychology are quantitative in nature. Quantitative researchers typically start with a focused research question or hypothesis, collect a small amount of numerical data from a large number of individuals, describe the resulting data using statistical techniques, and draw general conclusions about some large population. Although this method is by far the most common approach to conducting empirical research in psychology, there is an important alternative called qualitative research. Qualitative research originated in the disciplines of anthropology and sociology but is also used to study psychological topics. Qualitative researchers generally begin with a less focused research question, collect large amounts of relatively "unfiltered" data from a relatively small number of individuals, and describe their data using nonstatistical techniques, such as grounded theory, thematic analysis, critical discourse analysis, or interpretative phenomenological analysis. They are usually less concerned with drawing general conclusions about human behavior than with understanding in detail the experience of their research participants.

Consider, for example, a study by researcher Per Lindqvist and his colleagues, who wanted to learn how the families of teenage suicide victims cope with their loss (Lindqvist, Johansson, Karlsson, 2008). They did not have a specific research question or hypothesis, such as, What percentage of family members join suicide support groups? Instead, they wanted to understand the variety of reactions that families had, with a focus on what it is like from their perspectives. To address this question, they interviewed the families of 10 teenage suicide victims in their homes in rural Sweden. The interviews were relatively unstructured, beginning with a general request for the families to talk about the victim and ending with an invitation to talk about anything else that they wanted to tell the interviewer. One of the most important themes that emerged from these interviews was that even as life returned to "normal," the families continued to struggle with the question of why their loved one committed suicide. This struggle appeared to be especially difficult for families in which the suicide was most unexpected.

The Purpose of Qualitative Research

The strength of quantitative research is its ability to provide precise answers to specific research questions and to draw general conclusions about human behavior. This method is how we know that people have a strong tendency to obey authority figures, for example, and that female undergraduate students are not substantially more talkative than male undergraduate students. But while quantitative research is good at providing precise answers to specific research questions, it is not nearly as good at generating novel and interesting research questions. Likewise, while quantitative research is good at drawing general conclusions about human behavior, it is not nearly as good at providing detailed descriptions of the behavior of particular groups in particular situations. And quantitative research is not very good at communicating what it is actually like to be a member of a particular group in a particular situation.

But the relative weaknesses of quantitative research are the relative strengths of qualitative research. Qualitative research can help researchers to generate new and interesting research questions and hypotheses. The research of Lindqvist et al (2008), for example, suggests that there may be a general relationship between how unexpected a suicide is and how consumed the family is with trying to understand why the teen committed suicide. This relationship can now be explored using quantitative research. But it is unclear whether this question would have arisen at all without the researchers sitting down with the families and listening to what they themselves wanted to say about their experience. Qualitative research can also provide rich and detailed descriptions of human behavior in the real-world contexts in which it occurs. Among qualitative researchers, this depth is often referred to as "thick description" (Geertz, 1973). Similarly, qualitative research can convey a sense of what it is actually like to be a member of a particular group or in a particular situation—what qualitative researchers often refer to as the "lived experience" of the research participants. Lindqvist and colleagues, for example, describe how all the families spontaneously offered to show the interviewer the victim's bedroom or the place where the suicide occurred—revealing the importance of these physical locations to the families. It seems unlikely that a quantitative study would have discovered this detail.

Some contrasts between qualitative and quantitative research

Qualitative	Quantitative
In-depth information about relatively few people	Less depth information with larger samples
Conclusions are based on interpretations drawn by the investigator	Conclusions are based on statistical analyses
Global and exploratory	Specific and focused

Data Collection in Qualitative Research

Data collection approaches in qualitative research are quite varied and can involve naturalistic observation, participant observation, archival data, artwork, and many other things. But one of the most common approaches, especially for psychological research, is to conduct interviews. Interviews in qualitative research can be unstructured—consisting of a small number of general questions or prompts that allow participants to talk about what is of interest to them—or structured, where there is a strict script that the interviewer does not deviate from. Most interviews are in between the two and are called semi-structured interviews, where the researcher has a few consistent questions and can follow up by asking more detailed questions about the topics that come up. Such interviews can be lengthy and detailed, but they are usually conducted with a relatively small sample. The unstructured interview was the approach used by Lindqvist and colleagues in their research on the families of suicide victims because the researchers were aware that how much was disclosed about such a sensitive topic should be led by the families, not by the researchers.

Another approach used in qualitative research involves small groups of people who participate together in interviews focused on a particular topic or issue, known as focus groups. The interaction among participants in a focus group can sometimes bring out more information than can be learned in a one-on-one interview. The use of focus groups has become a standard technique in business and industry among those who want to understand consumer tastes and preferences. The content of all focus group interviews is usually recorded and transcribed to facilitate later analyses. However, we know from social psychology that group dynamics are often at play in any group, including focus groups, and it is useful to be aware of those possibilities. For example, the desire to be liked by others can lead participants to provide inaccurate answers that they believe will be perceived favorably by the other participants. The same may be said for personality characteristics. For example, highly extroverted participants can sometimes dominate discussions within focus groups.

Data Analysis in Qualitative Research

Although quantitative and qualitative research generally differ along several important dimensions (e.g., the specificity of the research question, the type of data collected), it is the method of data analysis that distinguishes them more clearly than anything else. To illustrate this idea, imagine a team of researchers that conducts a series of unstructured interviews with people recovering from alcohol use disorder to learn about the role of their religious faith in their recovery. Although this project sounds like qualitative research, imagine further that once they collect the data, they code the data in terms of how often each participant mentions God (or a "higher power"), and they then use descriptive and inferential statistics to find out whether those who mention God more often are more successful in abstaining from alcohol. Now it sounds like quantitative research. In other words, the quantitative-qualitative distinction depends more on what researchers do with the data they have collected than with why or how they collected the data.

But what does qualitative data analysis look like? Just as there are many ways to collect data in qualitative research, there are many ways to analyze data. One general approach called **grounded theory** (Glaser & Strauss, 1967) was developed within the field of sociology in the 1960s and has gradually gained popularity in psychology. In qualitative research using grounded theory, researchers start with the data and develop a theory or an interpretation that is *grounded in* those data. They do this analysis in stages. First, they identify ideas that are repeated throughout the data. Then they organize these ideas into a smaller number of broader themes. Finally, they write a theoretical narrative—an interpretation of the data in terms of the themes that they have identified. This theoretical narrative focuses on the subjective experience of the participants and is usually supported by many direct quotations from the participants themselves.

As an example, consider a study Abrams & Curran (2009), who used the grounded theory approach to study the experience of postpartum depression symptoms among low-income mothers. Their data were the result of unstructured interviews with 19 participants with the observed broad themes below in a table. In their research report, they provide numerous quotations from their participants, such as this one from "Destiny:"

> *Well, just recently my apartment was broken into and the fact that his Medicaid for some reason was canceled so a lot of things was happening within the last two weeks all at one time. So that in itself I don't want to say almost drove me mad but it put me in a funk....Like I really was depressed. (p. 357)*

Table of Themes and Repeating Ideas in a Study of Postpartum Depression Among Low-Income Mothers. Based on Research by Abrams and Curran (2009).

Theme	**Repeating ideas**
Ambivalence	"I wasn't prepared for this baby," "I didn't want to have any more children."
Caregiving overload	"Please stop crying" "I need a break" "I can't do this anymore."
Juggling	No time to breathe" "Everyone depends on me" "Navigating the maze"
Mothering alone	"I really don't have any help" "My baby has no father."
Real-life worry	"I don't have any money" "Will my baby be OK?" "It's not safe here"

Their theoretical narrative focused on the participants' experience of their symptoms, not as an abstract *affective disorder* but as closely tied to the daily struggle of raising children alone under often difficult circumstances.

The Quantitative-Qualitative Debate

Given their differences, it may come as no surprise that quantitative and qualitative research in psychology and related fields do not coexist in complete harmony. Some quantitative researchers criticize qualitative methods on the grounds that they lack objectivity, are difficult to evaluate in terms of reliability and validity, and do not allow generalization to people or situations other than those actually studied. At the same time, some qualitative researchers criticize quantitative methods on the grounds that they overlook the richness of human behavior and experience and instead answer simple questions about easily quantifiable variables.

However many researchers agree that the two approaches can and should be combined into what has come to be called mixed-methods research (Todd, Nerlich, McKeown, Clarke, 2004). One approach to combining quantitative and qualitative research is to use qualitative research for hypothesis generation and quantitative research for hypothesis testing. Again, while a qualitative study might suggest that families who experience an unexpected suicide have more difficulty resolving the question of why, a well-designed quantitative study could test a hypothesis by measuring these specific variables in a large sample. A second approach to combining quantitative and qualitative research is referred to as triangulation. The idea is to use both quantitative and qualitative methods simultaneously to study the same general questions and to compare the results. If the results of the quantitative and qualitative methods converge on the same general conclusion, they reinforce and enrich each other. If the results diverge, then they suggest an interesting new question: Why do the results diverge and how can they be reconciled?

Using qualitative research can often help clarify quantitative results via triangulation. Trenor, Yu, Waight, Zerda, and Sha (2008) investigated the experience of female engineering students at a university. In the first phase, female engineering students were asked to complete a survey, where they rated a number of their perceptions, including their sense of belonging. Their results were compared across the student ethnicities, and statistically, the various ethnic groups showed no differences in their ratings of their sense of belonging. One might look at that result and conclude that ethnicity does not have anything to do with one's sense of belonging. However, in the second phase, the authors also conducted interviews with the students, and in those interviews, many minority students reported how the diversity of cultures at the university enhanced their sense of belonging. Without the qualitative component, we might have drawn the wrong conclusion about the quantitative results. This example shows how qualitative and quantitative research work together to help us understand human behavior.

What Is Observational Research?

The term observational research is used to refer to several different types of non-experimental studies in which behavior is systematically observed and recorded. The goal of observational research is to describe a variable or set of variables. More generally, the goal is to obtain a snapshot of specific characteristics of an individual, group, or setting. As described previously, observational research is non-experimental because nothing is manipulated or controlled, and as such we cannot arrive at causal conclusions using this approach. The data that are collected in observational research studies are often qualitative in nature but they may also be quantitative or both (mixed-methods). There are several different types of observational methods that will be described below.

Naturalistic Observation

Naturalistic observation is an observational method that involves observing people's behavior in the environment in which it typically occurs. Thus, naturalistic observation is a type of field research (as opposed to a type of laboratory research). Jane Goodall's famous research on chimpanzees is a classic example of naturalistic observation. Dr. Goodall spent three decades observing chimpanzees in their natural environment in East Africa. She examined such things as chimpanzee's social structure, mating patterns, gender roles, family structure, and care of offspring by observing them in the wild. However, naturalistic observation could more simply involve observing shoppers in a grocery store, children on a school playground, or psychiatric inpatients in their wards. Researchers engaged in naturalistic observation usually make their observations as unobtrusively as possible so that participants are not aware that they are being studied. Such an approach is called disguised naturalistic observation. Ethically, this method is considered to be acceptable if the participants remain anonymous and the behavior occurs in a public setting where people would not normally have an expectation of privacy. Grocery shoppers putting items into their shopping carts, for example, are engaged in public behavior that is easily observable by store employees and other shoppers. For this reason, most researchers would consider it ethically acceptable to observe them for a study. On the other hand, one of the arguments against the ethicality of the naturalistic observation of "bathroom behavior" discussed earlier in the book is that people have a reasonable expectation of privacy even in a public restroom and that this expectation was violated.

In cases where it is not ethical or practical to conduct disguised naturalistic observation, researchers can conduct undisguised naturalistic observation where the participants are made aware of the researcher presence and monitoring of their behavior. However, one concern with undisguised naturalistic observation is reactivity. Reactivity refers to when a measure changes participants' behavior. In the case of undisguised naturalistic observation, the concern with reactivity is that when people know they are being observed and studied, they may act differently than they normally would. This type of reactivity is known as the Hawthorne effect. For instance, you may act much differently in a bar if you know that someone is observing you and recording your behaviors and this would invalidate the study. So disguised observation is less reactive and therefore can have higher validity because people are not aware that their behaviors are being observed and recorded. However, we now know that people often become used to being observed and with time they begin to behave naturally in the researcher's presence. In other words, over time people habituate to being observed. Think about reality shows like Big Brother or Survivor where people are constantly being observed and recorded. While they may be on their best behavior at first, in a fairly short amount of time they are flirting, having sex, wearing next to nothing, screaming at each other, and occasionally behaving in ways that are embarrassing.

Participant Observation

Another approach to data collection in observational research is **participant observation**. In participant observation, researchers become active participants in the group or situation they are studying. Participant observation is very similar to naturalistic observation in that it involves observing people's behavior in the environment in which it typically occurs. As with naturalistic observation, the data that are collected can include interviews (usually unstructured), notes based on their observations and interactions, documents, photographs, and other artifacts. The only difference between naturalistic observation and participant observation is that researchers engaged in participant observation become active members of the group or situations they are studying. The basic rationale for participant observation is that there may be important information that is only accessible to, or can be interpreted only by, someone who is an active participant in the group or situation. Like naturalistic observation, participant observation can be either disguised or undisguised. In disguised participant observation, the researchers pretend to be members of the social group they are observing and conceal their true identity as researchers.

In a famous example of disguised participant observation, Leon Festinger and his colleagues infiltrated a doomsday cult known as the Seekers, whose members believed that the apocalypse would occur on December 21, 1954. Interested in studying how members of the group would cope psychologically when the prophecy inevitably failed, they carefully recorded the events and reactions of the cult members in the days before and after the supposed end of the world. Unsurprisingly, the cult members did not give up their belief but instead convinced themselves that it was their faith and efforts that saved the world from destruction. Festinger and his colleagues later published a book about this experience, which they used to illustrate the theory of cognitive dissonance (Festinger, Riecken, Schachter, 1956).

In contrast with undisguised participant observation, the researchers become a part of the group they are studying and they disclose their true identity as researchers to the group under investigation. Once again there are important ethical issues to consider with disguised participant observation. First no informed consent can be obtained and second deception is being used. The researcher is deceiving the participants by intentionally withholding information about their motivations for being a part of the social group they are studying. But sometimes disguised participation is the only way to access a protective group (like a cult). Further, disguised participant observation is less prone to reactivity than undisguised participant observation.

Rosenhan's study (1973) of the experience of people in a psychiatric ward would be considered disguised participant observation because Rosenhan and his pseudopatients were admitted into psychiatric hospitals on the pretense of being patients so that they could observe the way that psychiatric patients are treated by staff. The staff and other patients were unaware of their true identities as researchers.

Another example of participant observation comes from a study by sociologist Amy Wilkins on a university-based religious organization that emphasized how happy its members were (Wilkins, 2008). Wilkins spent 12 months attending and participating in the group's meetings and social events, and she interviewed several group members. In her study, Wilkins identified several ways in which the group "enforced" happiness—for example, by continually talking about happiness, discouraging the expression of negative emotions, and using happiness as a way to distinguish themselves from other groups.

One of the primary benefits of participant observation is that the researchers are in a much better position to understand the viewpoint and experiences of the people they are studying when they are a part of the social group. The primary limitation with this approach is that the mere presence of the observer could affect the behavior of the people being observed. While this is also a concern with naturalistic observation, additional concerns arise when researchers become active members of the social group they are studying because that they may change the social dynamics and/or influence the behavior of the people they are studying. Similarly, if the researcher acts as a participant observer there can be concerns with biases resulting from developing relationships with the participants. Concretely, the researcher may become less objective resulting in more experimenter bias.

Structured Observation

Another observational method is **structured observation**. Here the investigator makes careful observations of one or more specific behaviors in a particular setting that is more structured than the settings used in naturalistic or participant observation. Often the setting in which the observations are made is not the natural setting. Instead, the researcher may observe people in the laboratory environment. Alternatively, the researcher may observe people in a natural setting (like a classroom setting) that they have structured some way, for instance by introducing some specific task participants are to engage in or by introducing a specific social situation or manipulation.

Structured observation is very similar to naturalistic observation and participant observation in that in all three cases researchers are observing naturally occurring behavior; however, the emphasis in structured observation is on gathering quantitative rather than qualitative data. Researchers using this approach are interested in a limited set of behaviors. This allows them to quantify the behaviors they are observing. In other words, structured observation is less global than naturalistic or participant observation because the researcher engaged in structured observations is interested in a small number of specific behaviors. Therefore, rather than recording everything that happens, the researcher only focuses on very specific behaviors of interest.

As an example, researchers Robert Kraut and Robert Johnston wanted to study bowlers' reactions to their shots, both when they were facing the pins and then when they turned toward their companions (Kraut & Johnston, 1979). But what "reactions" should they observe? Based on previous research and their own pilot testing, Kraut and Johnston created a list of reactions that included "closed smile," "open smile," "laugh," "neutral face," "look down," "look away," and "face cover" (covering one's face with one's hands). The observers committed this list to memory and then practiced by coding the reactions of bowlers who had been videotaped. During the actual study, the observers spoke into an audio recorder, describing the reactions they observed. Among the most interesting results of this study was that bowlers rarely smiled while they still faced the pins. They were much more likely to smile after they turned toward their companions, suggesting that smiling is not purely an expression of happiness but also a form of social communication.

In another example (this one in a laboratory environment), Dov Cohen and his colleagues had observers rate the emotional reactions of participants who had just been deliberately bumped and insulted by a confederate after they dropped off a completed questionnaire at the end of a hallway. The confederate was posing as someone who worked in the same building and who was frustrated by having to close a file drawer twice in order to permit the participants to walk past them (first to drop off the questionnaire at the end of the hallway and once again on their way back to the room where they believed the study they signed up for was taking place). The two observers were positioned at different ends of the hallway so that they could read the participants' body language and hear anything they might say. Interestingly, the researchers hypothesized that participants from the southern United States, which is one of several places in the world that has a "culture of honor," would react with more aggression than participants from the northern United States, a prediction that was in fact supported by the observational data (Cohen, Nisbett, Bowdle, Schwarz, 1996).

When the observations require a judgment on the part of the observers—as in the studies by Kraut and Johnston and Cohen and his colleagues—a process referred to as coding is typically required. Coding generally requires clearly defining a set of target behaviors. The observers then categorize participants individually in terms of which behavior they have engaged in and the number of times they engaged in each behavior. The observers might even record the duration of each behavior. The target behaviors must be defined in such a way that guides different observers to code them in the same way. Researchers are expected to demonstrate the interrater reliability of their coding procedure by having multiple raters code the same behaviors independently and then showing that the different observers are in close agreement. Kraut and Johnston, for example, video recorded a subset of their participants' reactions and had two observers independently code them. The two observers showed that they agreed on the reactions that were exhibited 97% of the time, indicating good interrater reliability.

One of the primary benefits of structured observation is that it is far more efficient than naturalistic and participant observation. Since the researchers are focused on specific behaviors this reduces time and expense. Also, often times the environment is structured to encourage the behaviors of interest which again means that researchers do not have to invest as much time in waiting for the behaviors of interest to naturally occur. Finally, researchers using this approach can clearly exert greater control over the environment. However, when researchers exert more control over the environment it may make the environment less natural which decreases external validity. It is less clear for instance whether structured observations made in a laboratory environment will generalize to a real-world environment. Furthermore, since researchers engaged in structured observation are often not disguised there may be more concerns with reactivity.

Quantifying Qualitative Data

From any qualitative source, from observations or interviews, behaviors can be coded, scored and counted. This process essentially turns the data obtained in a nominally qualitative method into quantitative data that can be used to support statistical analysis using the tools we have covered previously in this text. Doing so requires establishing a coding system targeting the behavior in question and using that to calculate a quantitative measure of that behavior, even something as simple as counting occurrences.

For example, a researcher studying relationship satisfaction among married couples might observe in open-ended interviews that couples having difficulty tend to show non-verbal expressions of contempt, such as eye-rolling, during discussions of dispute. This observational research might lead to the hypothesis that the rate of these expressions of contempt is correlated with scores of overall marriage satisfaction. Frequencies of these kinds of expressions could then be counted from recorded interactions and correlated with scores of relationship satisfaction. This would allow a quantitative evaluation of the strength of the association between these two measures. Of course, this would still be non-experimental research as neither of the variables evaluated were manipulated directly by the experimenter.

Qualitative research often focuses on observations in very unstructured contexts to generate hypotheses. This approach may be followed by other non-experimental approaches based on more qualitative measures related to the initial observations. Because of the challenges inherent to drawing conclusions from non-experimental studies, these studies might be followed up by experimental methods with carefully controlled conditions and random assignment to levels of a manipulated independent variable. Note that at each step along this kind of research plan, we might lose some aspects of external validity in exchange for greater internal validity and confidence in the conclusions obtained.

Key Takeaways

- Qualitative research is an important alternative to quantitative research in psychology. It generally involves asking broader research questions, collecting more detailed data (e.g., interviews), and using non-statistical analyses.
- Many researchers conceptualize quantitative and qualitative research as complementary and advocate combining them. For example, qualitative research can be used to generate hypotheses and quantitative research to test them.
- There are several different approaches to observational research including naturalistic observation, participant observation, structured observation, case studies, and archival research.
- Naturalistic observation is used to observe people in their natural setting; participant observation involves becoming an active member of the group being observed; structured observation involves coding a small number of behaviors in a quantitative manner; case studies are typically used to collect in-depth information on a single individual; and archival research involves analyzing existing data.

Exercises

Question 1. Qualitative research is often used in a research program as a first or preliminary study before research with experimental methods and manipulated variables. Why is this?

Question 2. Why might we use a qualitative research technique to study parent-child interactions and aggressive behavior in preschoolers? Outline what such a study might look like.

Question 3. In a diary study of conflict resolution in relationships, the research protocol could involve daily documentation of any conflict experiences and how they were or were not resolved. What kinds of events might be coded for and counted in a study of this type?

18 Ethics 2: RCR

In 1998 a medical journal called The Lancet published an article of interest to many psychologists. The researchers claimed to have shown a statistical relationship between receiving the combined measles, mumps, and rubella (MMR) vaccine and the development of autism—suggesting furthermore that the vaccine might even cause autism. One result of this report was that many parents decided not to have their children vaccinated, which of course put them at higher risk for measles, mumps, and rubella. However, follow-up studies by other researchers consistently failed to find a statistical relationship between the MMR vaccine and autism—and it is widely accepted now in the scientific community that there is no relationship. In addition, several more serious problems with the original research were uncovered. Among them were that the lead researcher stood to gain financially from his conclusions because he had patented a competing measles vaccine. He had also used biased methods to select and test his research participants and had used unapproved and medically unnecessary procedures on them. In 2010 The Lancet retracted the article, and the lead researcher's right to practice medicine was revoked (Burns, 2010).

However, the idea that there were scientific concerns about the safety of vaccines persisted in popular understanding and became a key part of skepticism about new mRNA vaccines that protect against the COVID-19 virus. Millions of people subsequently refused to use the vaccine and as a runs many thousands of people died unnecessarily. The damage caused by the original unethical report is likely to be higher than any other instance of scientific fraud in human history.

In the first chapter on research ethics, the focus was how to carry out scientific research in a manner that follows current expectations for best practice. We might summarize the core idea as “be nice to research participants.” Treat them with respect and design research to have value for the world (beneficence) and to make these benefits broadly available (justice). These ideas are then reflected in research processes related to obtaining informed consent and working with the Institutional Review Board as an external monitor of regulatory compliance with best practices.

The area described as **Responsible Conduct of Research (RCR)** reflects carrying out the scientific process in a fair and ethical manner related to the integrity of research and fairness in assigning credit to the researchers involved in the research process. The capitalization of the term is due to fairly recent changes in training of scientists requiring explicit engagement with these issues, especially at the student level (graduate students and undergraduate researchers). The focus on these specific topics has been driven by funding agencies (NIH, NSF) who seek to improve the reliability and quality of the research process.

Responsible Conduct of Research

The core ideas can be expressed simply: Don't Lie, Cheat or Steal. The main topic of RCR is to detail how these kindergarten ethics ideas apply to research processes in psychological science. Compliance with RCR principles is aimed to maintain the highest level of integrity in research processes so that the scientific community can rely on and trust the results of our scientific work. We have previously discussed the problem of the Type 1 error, a false claim of an effect among variables in research that turns out to be inaccurate. This can happen due to poor design or unexpected problems with extraneous variables. It can also happen due to an integrity violation where researchers do not follow best RCR practices. False claims that affect people's behavior based on the study can have substantial negative effects on society, such as with the example of vaccines. Unfortunately, identifying and attempting tor retract false claims appears to have a side effect of reducing non-scientists confidence in science overall, which can also lead to massive societal costs, for example due to skepticism about climate change.

Part of the motivation to increase awareness and training in RCR was due to the acknowledgment of external pressures on scientists engaged with the research processes. Successful science can produce substantial rewards for researchers including employment, promotion and access to research funding. Carrying out an unsuccessful research project is then costly in both time and opportunity loss that could reduce access to these rewards. The main content of this text is to illustrate the methods for carrying out research with the most rigor and care possible, but under external pressures, some researchers have failed to adhere to rigorous methods leading to results that are incorrect and/or retracted.

RCR Failure: Fraud

An obvious way that integrity can be violated in research practice is the wholesale fabrication of data, which is then presented as if it were properly collected. This is fairly rare, although not unheard of, in scientific work as the majority of modern scientific work is done in teams of researchers. This type of violation is so blatant and obvious that it is unlikely not to be known to other members of a collaborative research team. Purely fraudulent findings are also unlikely to be useful as part of a research program to drive subsequent research, which is often an important part of the general operation of a research group. Being found responsible for research fraud also effectively ends a scientist's research career, leading to immediate dismissal from the university or institution at which they work and a future bar on any external funding support.

RCR Failure: Falsification

A more pernicious issue in maintaining rigor in research processes relates to **falsification** of research findings, which covers a range of inappropriate data handling methods that lead to presentation of a false conclusion. The simplest of these is to exclude data collected that does not support the researcher's hypothesis. If we hypothesize that experimental condition A leads to higher scores on our DV than condition B, we can simply exclude all the low scorers from condition A (or high scores from B) and obtain an apparently statistically reliable result. This "data selection" is an obvious failure of research integrity and is treated in the same manner as wholesale fabrication.

There are subtler ways that aspects of falsification can creep into research. Scoring of subjective ratings of performance might not be done in a completely blind manner. Performance hints or clues could be given to participants in one condition. Bias could be covertly embedded in the task instructions or context. Participant recruiting could embed bias in assignment to conditions if not done properly randomly. These subtler issues are seen as problematic due to being difficult or nearly impossible to detect in the report of a completed research project. They could even be created accidentally by researchers who are simply so focused on research success that they implicitly deviate from best practice. A goal of RCR training for all lab personnel is to make everybody on the research team aware of these potential failures to provide checks on both their own work and the work of their collaborators.

The process of publishing research depends on peer-review of the methods, results and conclusions of a research project. Unfortunately, data handling problems are not visible to a peer reviewer, so this process does not effectively protect against RCR problems. In many modern journals, authors are encouraged or even required to publish their dataset in a publicly available location to support their results. However, even this may not protect against data selection if the publicly available data has already had the inconsistent data excluded.

A university or large research institution will typically have an office charged with evaluating scientific processes to assure compliance with best practices in research integrity, an **Office of Research Integrity**. If a concern is raised about a specific researcher or team, this office is charged with investigating and determining if an integrity violation has occurred. This investigation generally takes the form of an audit of research practices, review of raw data, preliminary analysis and evaluation of as many steps of the core research process as possible. These investigations can be complex and time-consuming as active research processes among a team of collaborators can often be fast-moving and sometimes important decisions are made quickly without immediate realization of how consequential they are.

Various recommendations have been put forth to improve the general process of research to maintain the highest levels of compliance with best practice. Some of these involve slowing the pace of research. For example, **pre-registration** of all research studies by reporting methodology, recruiting and planned analysis in advance of formal data collection. These obviously improve integrity but unfortunately can actually exacerbate the problem of external pressure to produce successful results by putting methodologically rigorous labs at a disadvantage in competition with labs that move faster and less carefully. Another approach is to document and record as much of the research process as possible so that if a question about integrity is raised later, an audit can verify if a problem occurs. One way to accomplish is to work as if there were cameras recording every aspect of the research process in the laboratory so that everybody possibly biased decision about participant exclusion, assignment to conditions or scoring could be evaluated later.

RCR Failures: Plagiarism and Research Privacy

One of the challenges with maintaining a robust record of all research practices is that some of the external pressure on scientists comes from competition to obtain an important discovery first. This problem is most evident in research fields like drug discovery where establishing the effectiveness of a new pharmacological agent can produce a patentable discover worth as much as a billion dollars. Large financial rewards for discoveries are exceedingly rare in psychological science but being the first to discover, name or characterize a novel aspect of psychology can be very rewarding in career advancement and scientific fame. As a result, cases do occur where multiple labs are considering very similar hypotheses and essentially racing to complete and publish their research project first. The lab that wins the race will accomplish a high-impact publication and lasting credit for the idea whereas the lab that finishes second will lament having gotten scooped.

For research being carried out in a context of this kind of competition, researchers will often work with a high degree of privacy about the research being carried out. Unfortunately, protecting research methods by operating in secrecy does not generally support the ability to provide oversight of those research practices. Thus, the areas with the most pressure to produce are also often the most difficult to verify. This is a difficult problem to solve, and the most common current approach depends entirely on training of all research staff.

This issue is where the concept of **plagiarism** plays a more visible role in ethical research practice. The most common form of the issue of plagiarism familiar to students has to do with copying another's words or ideas and claiming them as your own. This is rare in scientific publication as the majority of the scientific record is easily available to all. In addition, best practices in science are to thorough review background research and cite the relevant research to support the latest findings. Once findings are published, the problem of claiming credit for another researchers' idea is rare.

However, before publication of a novel finding, there is a risk of another research group finding out about a novel methodological approach to research and then appropriating this idea for their own without credit. This is an obvious integrity violation that can be difficult to deal with.

A particularly famous example historically is the famous work of Francis Crick and James Watson to identify the double-helix structure of the DNA molecule. They had been working on this problem for some time, as had several other large research labs, all of which were working secretly in order to be first to solve the problem. Crick and Watson were also serving as reviews for research grants and in that role saw preliminary data obtained by Rosalind Franklin using x-ray crystallography of DNA that was consistent with the idea of a double helix. That directed their subsequent work to show that DNA was constructed that way, which led to substantial scientific fame and a Nobel prize. It was not until decades later that Franklin's contribution to this discovery was fully appreciated and properly acknowledged. The ethical issues in this case are quite complex as there is no doubt that Crick and Watson developed completely novel methods and tools to come to their conclusion. When their Nobel prize was awarded, Franklin had passed away due to an unfortunately young case of ovarian cancer and since that award is not given posthumously, she would not likely have been included in any event. There are reports that she was even offered authorship on the original paper but declined (her own findings were published simultaneously).

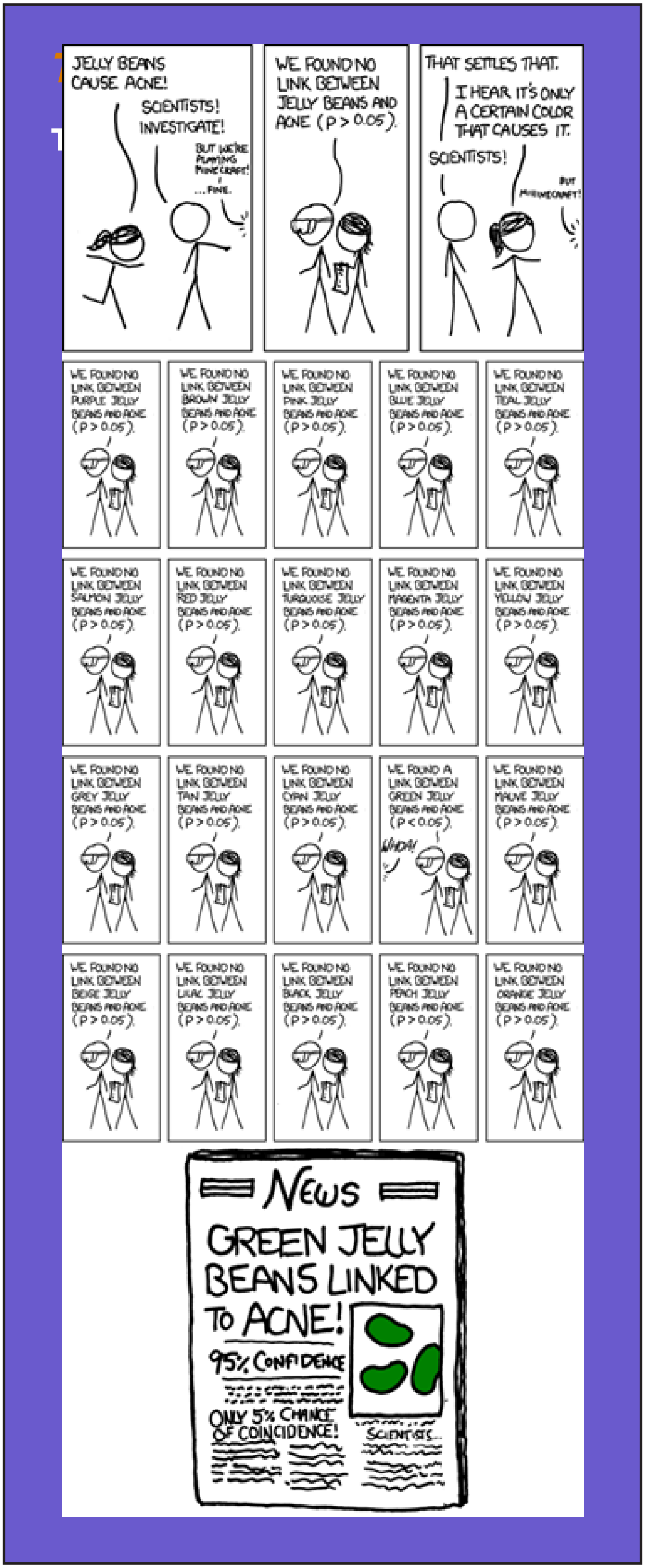
However, it is also clear that use of pre-publication data from another research lab without their knowledge or permission is clearly not following best practice for ethical research.

RCR: Authorship

Credit for scientific findings is generally reflected in participating in authoring the scientific report. Surprisingly, the rules for who is officially an author on a published report are not completely clear and consistent across all domains of science, or even all subdomains within psychology. The APA provides guidance that researchers who provide **substantial intellectual contribution** to a project should receive authorial credit. However, *substantial intellectual contribution* is not defined. Cases where a key idea is appropriated from another research group and used without credit are clear violations of this policy.

More complicated are questions about a research result that comes out of a lab with a number of staff members, graduate students and/or undergraduate student researchers. Some levels of participation in the project are considered to merit co-authorship, but in other cases, there might merely be a mention in the Acknowledgments section of the published paper. There are no fixed rules for how this is decided, and guidelines vary across individual laboratories and also across different subdomains of research.

Even trickier is the question of who takes on the coveted role of lead or first author of a manuscript reporting the results of a project done by a collaborating group. Again, there are no fixed rules for this, but it is generally up to the leader of the research team to set a clear policy that is known to all research personnel in advance and to follow that policy without bias or favoritism. The research team leader is known as the PI for Principal Investigator (almost always by the acronym to avoid the common misuse of the homonym *principle*). In most university laboratories, the PI is the professor who runs the lab, even if the research is in practice generally organized by a co-investigator such as a post-doctoral researcher (person with a Ph.D. but is hired by the PI not the university) or a junior faculty member. Difficult situations can emerge where there is interpersonal conflict among research team members, especially if a graduate student and supervising faculty member have differing opinions about credit. One goal of improving RCR guidelines is to avoid conflict around this issue and minimize or eliminate conditions where a lead researcher appropriates credit for work done by laboratory members



RCR: Analytic Flexibility

An issue that has gotten increased attention in modern scientific practice related to questions about reproducibility of research is the potential problems associated with a flexible approach to data analysis. In its more extreme forms, this produces the phenomenon known as **p-hacking** where a large number of analytic approaches are tried until one of them produces a result that meets the traditional criterion of $p<.05$. To understand the math behind this, remember that the p-value associated with an analysis is technically the probability of observing the data pattern under the null hypothesis. That is the same as saying that if there was no effect and all the DV measurements will random numbers drawn from the same distribution, what are the odds it would accidentally look like there was an effect? The standard criterion of .05 is effectively a 1:20 odds measure. So, if we had a true null effect but ran a study 20 times, we might accidentally observe what looks like a real result. If you are familiar with conditional probability math, you'll know that it's actually only about 65% like to get a false positive in this case, but the broader point that multiple rounds of data collection or multiple analysis techniques essentially mean that we are not effectively meeting the standard criterion.

There are statistical tools for correctly adjusting complex analysis of data where a large number of multiple comparisons are being considered. An example of this is where we might have a large number of measures on a large-scale survey and search for all possible relationships among all the variables. The conjoint probability of a false finding across tests is no longer .05 and tools such as **Bonferroni correction** should be applied to maintain rigor. The integrity concern arises from the fact that it cannot be easily determined from a published finding if the authors attempted a wide range of very different types of analysis but only published the one that produced an effect. A related phenomenon informally called the file-drawer effect is where a research group might have attempted a study multiple times and discarded all the data from null findings and selected only an unusual successful version of the experiment to present.

The real problem with issues related to analytic flexibility is that the normal process of experimental research often requires researchers to approach their project with a certain amount of creativity and flexibility. In particularly novel research, it can take several attempts to construct the best operational definition of a tricky construct, which will produce a series of unsuccessful experiments before a successful one. In other cases, the DV will not exhibit the exact effect predicted but another approach appears to show a robust finding (e.g., a recoding of the DV, or a related measure). As a result, researchers working with a high degree of rigor and integrity can still end up questioning their own processes as to whether they have accidentally biased their results through flexible procedures.

The idea of pre-registering experimental design and analysis plans is aimed at avoiding false positive reporting through analytic flexibility. This is highly effective, but has the unfortunate side effect of potentially eliminating report of interesting but unexpected findings that would be identified through more creative methods. Where possible, the best practice is to internally replicate an unexpected finding, that is, publish a report with multiple studies showing a consistent pattern. If flexibility was required for the first finding but it is a true effect, it should replicate in subsequent research. This approach is not easily applied in all areas of psychological science, however. Studies with expensive imaging methods, based on limited patient samples, or unique large-scale population surveys cannot be quickly or easily repeated. In those cases, considered analysis plans are important for the researchers and in addition, interested readers of the reports should likely look for parallel similar findings or subsequent research replications.

RCR: Conflict of Interest

In standard reviews of RCR-related issues, the potential for problems related to **conflict of interest** is also covered. This is relatively rare in experimental psychology research but is a general concern in areas such as health, education, marketing, and other applied research domains. The core of the idea is simple. If a physician is contracted to carry out part of an efficacy study on a new pharmacological agent that is potentially worth a billion dollars if the drug reliable works, that research is being done under extreme external pressure. If the lead researcher has a financial interest in the company making the drug, great care needs to be taken to avoid any possible implicit bias in research procedures that might undercut the validity of the findings. In these cases, extremely rigorous double-blind procedures are used with active external oversight to guarantee integrity. And unfortunately, there are still cases found later where these processes failed.

Most psychological science research does not assess hypotheses with large financial implications. These studies do sometimes have implications for political policies, educational practices or some health interventions. In particular, this can occur when research funding is provided by a private foundation that seeks to advance a specific idea or agenda. To maintain integrity, researchers are required to disclose all financial support to university oversight and to the journal editors when research is published. This allows identification of possible conflicts of interest where, for example, research funding was contingent on a successful finding, which is not a good circumstance to foster best practices for integrity.

Tension between Ethics and Science

Now that the basic structure of best practice for ethical research (Chapter 8) and ethical conduct of science (above) has been reviewed, it is useful to note a basic tension between best practices in research ethics and maximizing the internal validity of experimental research. It may already be clear that ethical conflict in psychological research is unavoidable. Because there is little, if any, psychological research that is completely risk-free, there will almost always be a conflict between risks and benefits. Research that is beneficial to one group (e.g., the scientific community) can be harmful to another (e.g., the research participants), creating especially difficult tradeoffs. We have also seen that being completely truthful with research participants can make it difficult or impossible to conduct scientifically valid studies on important questions.

Of course, many ethical conflicts are fairly easy to resolve. Nearly everyone would agree that deceiving research participants and then subjecting them to physical harm would not be justified by filling a small gap in the research literature. But many ethical conflicts are not easy to resolve, and competent and well-meaning researchers can disagree about how to resolve them. Consider, for example, an actual study on "personal space" conducted in a public men's room (Middlemist, Knowles, & Matter, 1976). The researchers secretly observed their participants to see whether it took them longer to begin urinating when there was another man (a confederate of the researchers) at a nearby urinal. While some critics found this to be an unjustified assault on human dignity (Koocher, 1977), the researchers had carefully considered the ethical conflicts, resolved them as best they could, and concluded that the benefits of the research outweighed the risks (Middlemist, Knowles, & Matter, 1977). For example, they had interviewed some preliminary participants and found that none of them was bothered by the fact that they had been observed.

The point here is that although it may not be possible to eliminate ethical conflict completely, it is possible to deal with it in responsible and constructive ways. In general, this means thoroughly and carefully thinking through the ethical issues that are raised, minimizing the risks, and weighing the risks against the benefits. It also means being able to explain one's ethical decisions to others, seeking feedback on them, and ultimately taking responsibility for them.

Several of the sources of risks to participants are both very common and very mild. Keeping participants blind to the experimental conditions has an element of deception built into the design. Maintaining a signed consent form and records of compensation (payment) to participants creates a privacy risk that somebody may discover they participated in the research study. The time spent completing the procedure is an inconvenience and even survey completion or simple cognitive tasks can be either cognitively challenging or boring. For these common design elements, providing advance information through the informed consent process and fairly compensating participants for their time addresses these mild issues.

However, it is worth acknowledging areas where the best experimental research practices become impossible due to ethical concerns. A number of these arise from research based on **interventions** aimed to provide a benefit to the participant as well as advancing science. A clear example is research aimed to establish the effectiveness of life-saving medical interventions. When research on drugs to treat and cure AIDS were being developed, it quickly became clear that while the best research design is a double-blind randomized clinical trial using a placebo, that meant condemning the control group participants to poor health outcomes or even death. The same is true for treatments of life-threatening cancer.

In these cases, it was decided that control groups would not be included in research designs and the treatment condition, which would be all participants, would be compared to population based outcomes of the disease. This is a weaker research practice because the sampling may well be biased depending on availability of the clinical trial and who enrolls is being compared with a broader population sample. However, it was decided that the ethical issues created for researchers and participants if there was true random assignment to control conditions far exceeded the added scientific value. And it might also be noted that maximally effective research practices are primarily needed to detect subtle effect sizes. If we optimistically hope that these medical treatments are having a large effect on health improvements or preventing death, a technically weaker research approach is still enough to establish efficacy.

The same concern about access to a benefit comes up in other areas of public policy research where social or educational interventions are deployed to help a segment of the population. In order to show a causal relationship of the intervention to improved outcomes, the intervention needs to be unavailable to a control group who do not benefit. For interventions that are expensive to implement, the need to show efficacy results in a need to be slightly unfair to participants in the control group. A standard practice in these studies is to make the intervention available to the control group after assessing the effectiveness of the intervention.

The problem of assignment to the control group was also a persistent issue with scientific research throughout the COVID-19 pandemic. The ability to truly establish the quantitative effectiveness of mask wearing on slowing the spread of the virus would have required random assignment of participants to non-mask-wearing conditions, which might have been life threatening. Instead, research in this area depended on non-experimental observations across mask wearing conditions that occurred based on personal choice or local culture. This led to widely varying estimates of mask effectiveness and criticism of the scientists attempting to study this. The problem was compounded by the fact that well-done science is not necessarily a fast process and there were demands for a more rapid answer. In addition, the estimates of mask effectiveness indicated it was not a huge effect at the individual level, although small effects over a large number of people showed robust positive epidemiological effects once that data could be collected.

HIPAA and PHI

Research related to COVID epidemiology also exposed an aspect of health-based research familiar to scientists who work in medical contexts (e.g., Neuropsychology, Chapter 20) known as HIPAA. **HIPAA stands for the Health Insurance Portability and Accountability Act**, a law passed to protect research participants from negative consequences associated with potential loss of privacy in health research. The insurance related issue was the fact that health insurance was once provided by companies who could deny coverage based on health status, known as pre-existing conditions. As a consequence, if you were in a clinical trial testing an AIDS treatment, you could potentially lose your health insurance for either having or being at risk for AIDS. HIPAA added oversight to handling of **Protected Health Information** (PHI) to improve the ethical practice of health-based science which affects a great deal of medical research and some psychological studies.

In the normal course of research, scientists are expected to protect the privacy of participants as much as possible but in standard practice this can not be done perfectly, especially with respect to the fact of participation. Participants might be seen coming to the research laboratory by others. While approved research staff are charged with carrying out research procedures, other people around the lab or research collaborators might become aware of details. Oversight of research practices with respect to informed consent or financial compensation methods may leave information about participation available. If research processes have need for keeping information about PHI, HIPAA applies to the research protocol and requires a series of improved information protection procedures aimed to reduce risk for participants. A full discussion of these methods is beyond the scope of this text but awareness of these special cases highlights more challenging elements of ethical research practice.

A common discussion/misconception related to the COVID pandemic might be useful as illustration. Asking people about their COVID vaccination status was in some cases thought to be a violation of HIPAA due to the need to reveal what would normally be PHI. However, self-disclosing PHI is not covered by HIPAA. It would only be relevant if someone were to access your medical records directly without your consent. Asking somebody about PHI might be rude, but it not technically an ethical violation. It should also be noted that ethical issues like this are never thought to have clear black-and-white answers in all cases. It would not be unreasonable for a decision to be made that prioritizes public health and the spread of disease over individual privacy. This kind of issue is also made complex by technological advance. At the time of this writing, it is not clear how ethical practice related to PHI/HIPAA is applied to online systems that give external proof of vaccine status that appears to derive directly from medical records. While these would seem to be in a gray area, they have been widely accepted and seen as valuable, likely meaning that the broader understanding of how PHI is handled will continue to evolve and regulatory guidelines will continue to be modified and improved.

Data Sharing

For sensitive data like PHI, researchers need to work carefully within guidelines in order to share data with collaborators. In general, data sharing is done by first de-identifying data, which is to remove all information in the data records that would link performance data back to the specific participant in the study. In many cases, this is as easy as coding data by participant id and avoiding the use of name (or email) in data records. If data cannot be effectively de-identified, then a research sharing agreement is written and reviewed by the IRB to evaluate any risks associated with possible privacy exposure.

Identification of participants is another area where technological advances have led to changes in oversight procedures. Many years ago, researchers might freely share biological specimens from human participants research with other labs. However, the advent of DNA sequencing means that blood or tissue samples can be analyzed in a way that reveals the original participant and now must be evaluated for privacy risk. It has been suggested that the same aspect may be true of some neuroimaging data, i.e., that brain images might be uniquely identifying eventually even though the tools for this do not yet exist. Machine learning techniques may also somebody be able to recover identity from extensive survey data if enough relatively individual data has been collected. In most cases, the privacy risk is minimal but data sharing should always be handled carefully to maintain compliance with best practice.

Waiver of consent

In some field research studies, research practice requires data collection without the ability to first provide informed consent. This can be approved by the IRB through a request for a **waiver of consent**. This process is also used in cases of severe privacy risk, such as data collection about criminal or high-risk behavior. If participants are potentially asked about prior behavior that could have immediate legal consequences, carrying out the research requires absolute privacy protection. In this case, the written consent process can be waived to ensure that no method of tracing the data back to the participant exists. This is obviously a highly specialized case and one that it very carefully and extensively reviewed. One aspect of the complexity of this process is that although ethical practice might require protection of all research records, it cannot be guaranteed that a legal process such as a court order might override institutional preferences. Ethical practice guidelines from the IRB, the university or even federal funding agencies do not have legal standing to stop a court order or warrant.

Mandatory Reporting

A fairly recent example of tension between legal understanding of ethical practice and science is recent decisions requiring some research personnel to be **mandatory reporters** for some kinds of observations. The most common situation for this is in developmental research with young children. If research personnel suspect that child abuse is occurring, they are required to report this to institutional authority for investigation. This appears to be a privacy violation for the child and their family, yet the decision was made that the need to protect children overrides the privacy concern. This policy has additional consequences such as the need for training of research staff in these laboratories to appropriately decide when reporting of suspicion is required.

A related issue arises in studies of mental health among adolescents, even university students. Measures related to clinical depression have to be used carefully in practice as there may be institutional or legal policies in place requiring intervention for adolescents at risk. If answers to a mental health survey indicate a potential for self-harm, it may be necessary to have trained and qualified mental health professionals available for immediate participant support. A counter-intuitive consequence of this policy is that questions related to high levels of risk (e.g., self-harm) are often practically removed from mental health surveys if the research team does not have access to adequate support services. That is, knowing that a participant is at risk and not acting is deemed to be a worse situation than not asking questions that would identify the risk. This is clearly an example where scientific practice and ethics have fallen into a complex gray area without an obvious solution.

Summary

In most psychological research, effective compliance with best practices for ethical research are straightforward and easy to carry out. Showing respect for persons, ensuring voluntary cooperation, properly protecting privacy and data are usually easy to implement. Specialized procedures and training become critical for working in some select, more challenging subdomains. Awareness of the more complex issues in these domains is sometimes useful for understanding the long lists of questions asked in the process of obtaining approval for research from the IRB. Their processes have to be effective for all research and allow for determination of what the risks levels are without any assumptions about the usual case of minimal risk associated with much psychological science.

Key takeaways

- Everybody on the research team needs to be informed about standard practices and policies related to the responsible conduct of research to ensure fair and accurate scientific processes
- Data need to be handled with extreme care to avoid any possibilities of bias in selection that would weaken inferences from analysis
- Citation of others' influencing research is standard and appropriate. Use or reference to others' unpublished work should not be done.
- Authorship rules for publication vary across laboratories but generally adhere to a policy of authorship to individuals providing a significant intellectual contribution.
- Any information about participants related to personal or private health information must be handled carefully.
- Intervention research poses additional challenges to ethical science and is done in close collaboration with the IRB

Exercises

Question 1. Suppose a friend doing a research project said to you, "I'm sure my hypothesis is correct, so I'll just give my participants a hint here and there to make sure the data come out properly."

1a. What kind of RCR violation is this?

1b. What methodological approach should be used by this research to avoid this problem?

Question 2. In a study of marriage relationships, a researcher discovers that one of the participants in the study is an acquaintance and has indicated 'yes' to the question of having had an extra-marital affair. The researcher is debating whether they should inform the participants' spouse.

Question 3. A researcher doing a study on academic performance of students who have been diagnosed with ADHD. Some of the potential participants have refused to answer the question about their diagnosis, so the researcher contacts their medical provides without the participant's knowledge. What kind of research ethics violation is being considered here?

Question 4. Give an example of a research study not included in the chapter or which it would be unethical to include a placebo/control group.