Measurement

Quantitative research methods are used to create and test theories, but they use numbers to achieve this goal. Measurement in most psychology research involves attaching numerical values to observable phenomena, then working with the numerical values to determine differences, relationships, and patterns in the data. For example, a score on an IQ test is an observable indicator of intelligence, just as “number of times I wrecked dad’s car” is an observable measure of stupidity. Because psychologists measure so many different kinds of things, and because working with the numbers properly is so critically important, we have to look very closely at our measurement techniques and at what numbers really tell us. This chapter explains the properties of measurement scales and introduces some common problems we face in collecting data.

Properties of numbers

Different kinds of numbers can tell us different things. Numbers have four properties, and these properties determine how numbers can be used.

Identity: numbers can be used as identifying labels. For example, race cars are numbered so that spectators can identify them easily. When #3 flies by, you know who’s car it is and who’s in it. This use of numbers is not really different from using any other symbol, for example car #3 could just as well be called car #E because it’s value (3) is not numerically different from a car labeled “#2” in this context.

Magnitude: The more common use of numbers is to indicate a value. The magnitude property of numbers indicates that they reflect an amount of something, and larger numbers indicate more of this quality. Two heads are better than one, three's a crowd, four horses of the apocalypse are more than three horses, five fingers beats the Simpsons’ four, etc.

Equal intervals: Some numbers indicate amounts or magnitudes in such a way that the differences between numerical values mean the same thing all along the scale from low end to high end. For example, on a ruler, the distance between centimeter markings all along the ruler are the same throughout. On an attitude scale, the difference between a score of 20 and a score of 25 is the same, psychologically, as the difference between 25 and 30. The equal intervals property allows some types of math to be performed, such as adding and averaging.

True zero: Sometimes numbers can have a value of zero that really means something. On an attitude scale, getting a score of zero does not mean you have absolutely no amount of the attitude being assessed, or that you're no longer with us;
it just means that you didn’t answer any of the questions in a way that would give you any points. However, on a weight scale, getting a zero means that you’re not on the scale: there is no weight and you have disappeared. For qualities like length, weight, number of car accidents, and years in jail the value zero really means zero. The true zero property allows a full range of mathematical operations.

These four properties define the four kinds of numbers that we use in psychology, discussed in the next section.

Nominal Scales

A nominal scale is one that has only the identity property and is merely used as a label. For example, it is common in research to use numbers for categorical variables when the data are entered into a data analysis software package, e.g., males=1 and females=2.

Ordinal Scales

When the property “magnitude” is present in a scale, values on the scale can distinguish between lower and higher amounts of a psychological construct. Scales that have the magnitude property but not the equal-intervals property are termed ordinal scales. Ordinal scales indicate relative differences on a construct but are otherwise not very precise. The most common type of ordinal scale is a rank-order scale. For example, when you organize a classroom of children by height then assign them rank values “1,” “2,” “3,” etc., the scale is ordinal. If the tallest child were 6′2″ the second 5′6″ and the third 5′4″, the rankings would accurately reflect the relative differences but would not indicate how large the differences were. (See figure.)

Psychologists only use ordinal scales when they can’t do any better. For example, some approaches to studying values suggest that value questionnaires must require respondents to rank their values from most important to least important. No ties are allowed. The resulting data are on an ordinal scale. (See sidebar for a well-known questionnaire)

Post-Materialist Values (Inglehart)

Rank these values:
___ E. Maintaining order in the nation
___ F. Giving people more say in government decisions
___ G. Fighting rising prices
___ H. Protecting freedom of speech
___ A. Maintaining a high level of economic growth
___ B. Making sure this country has strong defense forces
___ C. Seeing that people have more to say about how things are done at their jobs and in their communities
___ D. Trying to make our cities and countryside more beautiful
___ I. A stable economy
___ J. Progress toward a less impersonal and more humane society
___ K. Progress toward a society in which ideas count more than money
___ L. The fight against crime

Scoring:
Add the ranks for items A B E G I L. Total = ____ (a)
Add the ranks for items F J C K H. Total = ____ (b)
Subtract (a) - (b) Total = ____

Interpretation: A high value means you are a “postmaterialist” or “postmodern” person: someone who is more concerned about quality and meaningfulness in life than meeting basic needs. Post-materialists tend to have experienced safety, peace and wealth during their teen years.
Ordinal data are difficult to analyze. Measurements in which the property “equal intervals” is added on top of “magnitude” are called interval scales. Because it has the equal intervals property, this sort of scale produces much better information than an ordinal scale. Most psychology tests yield interval scale data. For example, we assume that the difference between an IQ of 90 and an IQ of 95 is the same as the difference between an IQ of 95 and an IQ of 100. If we were to rate the values in the Post-Materialist Values questionnaire (sidebar) on 10 point scales ranging from 0=don’t agree at all to 10=completely agree, the resulting questionnaire score would be on an interval scale.

Psychologists prefer interval scales because most of the statistics used to analyze psychological data require data that are at least this good. Scores on an interval scale can be averaged, so we can calculate the average attitude in a group; they can be added and subtracted, so we can calculate the difference between scores on a depression scale before and after psychotherapy. Interval scales are sometimes called “score data.”

Interval Scales

Measurements in which the property “equal intervals” is added on top of “magnitude” are called interval scales. Because it has the equal intervals property, this sort of scale produces much better information than an ordinal scale. Most psychology tests yield interval scale data. For example, we assume that the difference between an IQ of 90 and an IQ of 95 is the same as the difference between an IQ of 95 and an IQ of 100. If we were to rate the values in the Post-Materialist Values questionnaire (sidebar) on 10 point scales ranging from 0=don’t agree at all to 10=completely agree, the resulting questionnaire score would be on an interval scale.

Psychologists prefer interval scales because most of the statistics used to analyze psychological data require data that are at least this good. Scores on an interval scale can be averaged, so we can calculate the average attitude in a group; they can be added and subtracted, so we can calculate the difference between scores on a depression scale before and after psychotherapy. Interval scales are sometimes called “score data.”

Ratio Scales

Adding the property “true zero” to an interval scale elevates its status to a ratio
The term ratio scale is used because numbers on such a scale can be used in mathematical operations such as ratios. For example, weight is on a ratio scale and we can speak of someone being twice as heavy as someone else, or half as tall. Without a true zero point this math cannot be done. On interval scales such as a 10-point values rating scale, a score of “8” does not indicate twice as much agreement with the value as a score of “4.” An IQ of 140 does not indicate twice as much intelligence as an IQ of 70.

Ratio data are used less commonly than interval level data in psychology because we don’t usually measure physical qualities. However, for the purposes of statistical data analysis, interval and ratio scale data are nearly equivalent in psychology research.

**Effective Range**

Interval and ratio level measurements must be tuned to pick up the whole range of possible values in whatever is measured. This means that the measurement scale has to extend down to the lowest possible value and up to the highest. It also means that the scale must be able to detect small differences if these differences are important to the research. For example, an IQ test that can only measure IQs from 90 to 110 will be appropriate for a large number of people, but will miss a lot of others. A self-esteem scale designed for college students (normal people) might be inappropriate for a sample of CEOs (enormously high self-esteem people) because the CEOs would all be close together, around the top of the college student scale (see figure). In such a case, the scale would not discriminate among the CEOs even though there may be some interesting differences among them. A CEO-esteem scale would need to amplify the differences within this narrow range to be effective.

**Scale Attenuation Effects**

When the effective range of a scale is too restricted, scale attenuation effects may take place. A ceiling effect occurs when the upper end of the scale is too low to accommodate the data. The arrogant CEO example above illustrates a ceiling effect because the CEOs crowded together near the top of the scale. The scale can be completely inadequate to assess the underlying concept because the responses...
are either all at the very top of the scale or are beyond what the scale can measure. For example:

Terrorism is bad:

Strongly disagree ____ ____ ____ ____ ____ Strongly agree

Most respondents would check the highest point on the scale, but the actual opinion is probably beyond “strongly agree.” The solution to this kind of problem is to modify the response scale to detect a larger range of opinions or to use better items, e.g., “Terrorism is as bad as war” or “Terrorism is like genocide.”

Ceiling and floor effects. The red line represents the underlying psychological construct, for example, attitude toward anchovies. Individuals have attitudes ranging from very negative at the left end to very positive at the right. Each arrow indicates how a person with a particular attitude would express the attitude on a scale. Scale 1 is not able to pick up the attitudes of anchovie lovers, and Scale 2 is not able to pick up the attitudes of anchovie haters.

A floor effect is the same concept occurring at the bottom of the scale. Floor effects can take place on the types of scales just described, but they also are common on measurements that involve counting events. For example, if you were interested in counting terrorism events and chose to define the scale as “number of terrorist attacks in Florida per month” you will (hopefully) have a floor effect because most months will yield a count of zero. However, if you defined the scale as “number of terrorist events in the world per year” you will find values above zero.