

 **SAGE** researchmethods

Project Planner

Data Analysis and Interpretation

Title: Data Analysis and Interpretation
Originally Published: 2017
Publishing Company: SAGE Publications, Inc.
City: London, United Kingdom
ISBN: 9781526408570
DOI: <https://dx.doi.org/10.4135/9781526408570>

Contact SAGE Publications at <http://www.sagepub.com>.

This PDF has been generated from SAGE Research Methods.

This stage will:

- Explain different types of qualitative data analysis
- Explain different types of quantitative data analysis
- Help you to interpret the results of your data analysis

Once you have decided on your method of data collection and have gathered all of the data you need, you need to decide how to analyze and interpret your data.

This can be done in a variety of ways, and will depend on whether you used qualitative, quantitative, or mixed methods of data collection.

How Do I Analyze and Interpret Qualitative Data?

The first thing to do is to read through all of your textual material, listen to all of your audio material, and look carefully at all of your visual material. You have to have a holistic sense of what you have and you have to form some *preliminary* impression as to what it all means. “Preliminary” is emphasized here because as you go through the materials in any systematic way, your first impressions must be changed if necessary.

We can think of organizing your analysis and interpretation in terms of:

- **Ways of measuring from your qualitative material:** This may be, and mostly is, nothing more complicated than sorting things into kinds so the measurement is nominal.
- **Establishing meaning:** We have “patterning” as one style and “understanding” as another. Of course we tend to do both as we work through the materials. Some formal methods have both approaches built into them.

You may organize your analysis using one of these methods. Or you may choose to use one or more in combination. However, when using multiple methods, careful consideration must be given to integrating the resulting disparate elements into an overall synthesis.

Classification involves sorting things into kinds. This can be a process of classifying and naming elements as a measurement device. Or it can involve the development of a system of concepts. It can be both.

Induction is a more or less formal development of hypotheses on the basis of elements of the material which are tested against other elements.

Content analysis involves elements in the material being counted in order to generate a set of numbers for statistical analyses. This can be very simple. For example, the number of times an issue is identified in a set of interviews can be counted and reported on as a frequency count. More complex statistical approaches are appropriate if there are a very large number of cases.

Qualitative comparative analysis (QCA) is the systematic description of cases, in terms of the presence or

absence of attributes, in order to construct a data set which can establish configurations in a truth table.

Event analysis is the identification of specific events in time. The objective is to identify the events, their duration, and their sequence.

Discourse analysis focuses on communication among sets of individuals. It looks at the content of communication and the patterns of interaction in communication. Conversation analysis focuses on the latter aspect.

Framework analysis is a method developed in health research contexts. It is particularly suitable when large research teams with diverse backgrounds are engaged in research. It works best with material which is relatively homogeneous and focused on specific topics.

Grounded theory is a systematic constant comparison, guided by theoretical sampling, toward exhausting the explanatory potentials of the material.

Hermeneutic interpretation is the systematic deployment of the hermeneutic method in order to identify meaning. **Phenomenological approaches** emphasize the meanings developed by individual actors in their own life worlds.

Narrative analysis focuses on the actual structure of accounts as they are presented in narrative form within the materials.

[Search for resources about qualitative data analysis](#)

What Qualitative Analysis Technique(s) Should I Use?

Here are the factors which should guide your decision. You should choose approaches:

- which generate answers to your own research questions
- which you are comfortable with and feel competent in using
- which can be employed using the resources that are available to you
- which are generally acceptable in your own discipline or field

OR

- where you can provide a convincing argument as to why this innovation should be acceptable in your own discipline or field.

What Is Narrative Analysis?

Transcript

00:00

[MUSIC PLAYING] Hello, welcome. Professor Corrine Squire is co-director of the Center for Narrative Research at the University of East London. She's also a member of the NCRM Novela Research Methods. Novela stands for narrative of everyday life and linked

00:22

approaches. Thank you for speaking with me in conversation with Katie Wexler from Sage Publication. So the first question is, could you start by briefly explaining what narrative research is? Sure. Narrative research involves researching with materials that have been produced, which is deliberately stories to tell you.

00:46

It also involves collecting their stories by actually asking people to tell stories that many people just collecting materials and then treating the stories, if you think that there's a good reason for doing so. It involves analyzing the materials and in terms of what are interesting features of history

01:07

and the kinds of materials that include spoken narratives, written narratives, vision narratives various. Sometimes narratives actually turned into a time when the narrative filled with objects. [MUSIC PLAYING]

01:36

What kind of research questions narrative especially appropriate for? What I think, it's particularly useful. It's all in a phenomenal way of trying to find out things like new, perhaps not spoken about very fully. Difficult sometimes to speak about

01:57

because with people do you notice that you give people some space time, that things are to develop more complex than with other forms of research. I think it also is very likely to move between the levels of material rather than just say for instance in my search just quite a few people

02:19

living with HIV you hear from them to pleasures of the two stories at the same time those are stories that go on why do cultural stories with HIV being addressed and that also draws political narratives.

02:50

Are there several different approaches to doing narrative research? How would I know which one to use? This can be confusing at the beginning. My suggestion to people starting out is, usually, that they read very widely and they try to argue out with themselves and others which

03:10

approach works best for them. I mean, you do have to steer your own path towards taking an approach. There are just lots of different levels. So she might be looking at narratives at the level of everyday conversation. The way that people start to exchange with each other. You might need a full biographical narrative.

03:32

People could use whole life stories, and such. And I was curious, you might want to look at cultural narratives now that exist within popular culture, or political discourse, and how they relate to the first pieces. There are lot of different levels of narrative material and narrative analysis.

04:02

[MUSIC PLAYING] How would you collect your data? Well, there are a number of ways to treat it. You may want to try to work with data that exists.

04:25

That is about now, that includes active here and naturally in some ways this could be written data and media data, or data naturally occurring conversation. You might be actually putting people in coffee shops, or if you go to the internet material. I received a phone, it was explicitly

04:45

to ask people to tell you that story. So you might imagine interviewing, directive, getting to tell stories for my group and asking for the whole life story. Alternatively, you might choose to work with material not just thought, but it appears in the course of regulars in the structured interview, for instance. Or unstructured interview, because normally it

05:07

gets people safe to respond in interviews about stories. What's the difference between narrative analysis

05:28

and other forms of the problem they could take an hour that, for example, critical discourse analysis? Some people are not very special quality of their own, in that it gives voice to people's expressions of

their own identities and sounds. I think more generally, say now as it involves

05:51

working with material, themes, seem to get developed sequentially in ways that they don't actually live and that you will be paying attention to or to discuss. I was looking at small seconds of just causes the process to the narrative.

06:12

And others can be seen, actually, in part just isn't. Some people do view it like that. I think it's cracked as less good than sometimes. I'm just looking at large scale.

06:39

If I were just starting out my narrative narrative research project, could you give me some key bit of advice? So, I think the first thing, which is probably the most boring, is just to read a lot. Discuss it with friends and people you're working with. The second thing would be to work with material that you've

07:00

gotten from your research and to do that again with other people, to try and test out ways of analyzing together, in groups. The third thing is something that relates to the advice that was given to me, which is do a lot of crossword puzzles. I think my advice, in terms of narrative research

07:22

would be to look at a lot of popular culture.

[Narrative analysis](#) is founded around the recognition of the central role that constructing narratives plays in human experience. Narratives are ways in which people articulate their accounts of experiences in the social world and the ways they communicate with each other.

In its simplest sense, narrative analysis is a qualitative method which takes narratives as the object of research. It requires attention to:

- The content of the narrative
- The form, the way in which the narrative is organized.

Narratives usually take the form of texts or textual transcriptions of oral accounts. However, narrative analysis can also be applied to images. Film and video generally have a narrative structure but we can also see still images as narratives.

There are more formal and structured approaches to narrative analysis, particularly in organizational studies,

but there is no single way of doing narrative inquiry.

Transcript

00:00

[MUSIC PLAYING] I and nobody else would want to deny that narrative is important, in a sense, in its own right.

00:20

There is, of course, a long tradition of narrative analysis deriving from linguistics, and socio-linguistics, from discourse analysis, from linguistic anthropology. And those are very important. And I want to come back to that in just a second. What I want to argue, however, is that too much of what's called qualitative research,

00:44

at the moment, actually doesn't pay due respect to narrative analysis. And doesn't treat narratives in the way that I would want to see them treated. What do I mean by that? Well, a lot of people do what they say is narrative analysis, through interviews, very often, or through other means of data collection.

01:06

They collect what they call narratives. And very often you say, well, what do they do with them then? And what they do is they reproduced them. They celebrate them. They argue, or at least imply, that by collecting these narratives, by reproducing these narratives, somehow

01:27

they're giving us, the reader or the hearer, access to the private, the personal experiences of the individual narrator. Well, of course that's important. But it's not the whole story. And when I say I think we need to rescue narrative

01:49

from that kind of qualitative research, what I'm saying is OK, we have narratives. Narratives are pervasive. People produce stories in their everyday lives. People produce stories in their work environments. People produce stories in order to create moral tales, in order to create tales of success or failure.

02:10

But what are they doing with those stories? Or more, how are they constructing them? Narratives are

not unmediated, direct representations of your experience or mine. They're always shaped. They have a form. They have a function. So as analysts or social scientists, we have to be, I think, careful not only

02:33

to pay attention to what people are talking about, but also how. How are they talking about it? How are they constructing the stories? And if we return to what I said earlier, return to the work of linguists, return to the work of discourse analysts, actually return to the work of some very

02:53

important literary theorists too, recognize that not only are these not transparent, they're not unmediated. They actually have very clear structures. They have clear conventions. We have repertoires of ways of constructing stories. We have types of stories, like genre of stories.

03:18

So having collected them, then we should analyze them. We should understand how they're put together as particular kinds of social action, what kinds of consequences one story has over another, or one way of telling a story over another. How do they construct the teller as a particular kind of person?

03:40

How do they construct them as a particular kind of moral actor? How do they construct their lives retrospectively and prospectively. So the argument, in a sense of why do we need to rescue narratives, isn't to downgrade narrative. It's not to ignore narratives. On the contrary, it's what, I think,

04:02

is to give them their full weight is to recognize that they are, themselves, social forms, social products, social conventions that are used in particular kinds of cultural and social media, done to organize, to reconstruct the biographical,

04:24

to construct experience, not simply to reproduce it unproblematically. And therefore, our job as social scientists-- and I don't have a problem with calling myself a social scientist-- is to treat those as data, like any data, analyze them, as we would analyze any other kind of data,

04:44

in terms of their organization, their indigenous, inherent structures, and in terms of the forms and functions that they perform. So I'm arguing for a much more, if you like, disciplined, systematic, sustained analysis of narrative, not just as a thing in it's own right,

05:08

as it were, not just to do narrative analysis, not to take these narratives and rip them out of social context, and then treat them as if they were a thing just in themselves. But to analyze them and study them within a cultural context, within an organizational setting, in the context of people's careers,

05:30

the organizations that people work in, to recognize that narratives are, themselves, part of work settings, a way of organizing work. Social workers tell narratives in order to construct cases for their fellow workers. Lawyers tell narratives in order to construct cases about their clients.

05:51

Medical practitioners construct narratives about their patients and about themselves to their patients, to their clients, to their fellow medical practitioners in constructing cases, in presenting cases at conferences, and so on. Scientists tell narratives about their discoveries.

06:11

We don't need to just celebrate them. We need to analyze how does a scientist account for a discovery. By telling a story in a particular way, by using particular kinds of narrative conventions in order to do modesty, in order to do boasting, in order to account for how their discovery was

06:32

a matter of chance, but also hard work, how they stumbled upon the solution, but in the context of being absolutely brilliant scientists, and so on. So what I'm saying, and therefore, just to repeat myself, is collect narratives in context, collect narratives as part of systematic, ethnographic

06:56

fieldwork about organization, settings, and so on. And then, treat them to sustained, systematic, disciplined analysis. And if you do that, you could say a great deal more about those narratives, than if you simply collect them, chop them up, produce a few extracts, and as it were, say there you are then.

07:18

[MUSIC PLAYING]

[Search for resources about narrative analysis](#)

What Is Discourse Analysis?

[Discourse analysis](#) is about how people use language in communication. As with narrative analysis, there is

a focus not just on content but on form. In [conversation analysis](#) the actual form of the discourse, in terms of stops, starts, clues for continuation etc., is the primary focus of attention.

Discourse is a term for communication between or among two or more people. Social life is managed through linguistics and other related interactions among people. Language, which may include non-verbal signs, is the general vehicle which carries such interactions.

The interest in discourse is a reflection of the general linguistic turn in philosophy, the humanities, and the social sciences. Language becomes a central or even the central focus of attention. All variants of discourse analysis focus not just on what is said or written or communicated in any way, but also on how this is done. So there is a considerable overlap with the logic of [narrative analysis](#).

- Foucault asserted the significance not just of discourse but of discourses—that is ways in which knowledge becomes formalized into patterns which are manifest both in the structures of formal disciplines and in sets of social practices, for example in the organization of systems dealing with madness.
- [Critical discourse](#) analysis originates in a radical turn in linguistics. It focuses on issues of authority and power and combines:
 - A micro-level attention to the text of the discourse itself
 - A meso-level attention to the context and processes of production of the text
 - A macro-level concern with the way in which social structures, hierarchy, and related processes shape both the text at the micro-level and all the processes which surround the meso-level of its production.

[Search for resources about discourse analysis](#)

[Search for resources about critical discourse analysis](#)

What Is Conversation Analysis?

Where discourse analysis in general looks at any form of discourse, [conversation analysis \(CA\)](#) is concerned solely with the actual processes of interpersonal communication in social interactions between individuals and does not include written texts or larger socio-cultural phenomena.

Conversation analysis is an approach to the study of social interaction, both verbal and non-verbal, in situations of everyday life. As its name implies, CA began with a focus on casual conversation, but its methods were subsequently adapted to embrace more task- and institution-centered interactions, such as those occurring in doctors' offices, courts, law enforcement, helplines, educational settings, and the mass media.

CA is an established method used in sociology, anthropology, linguistics, speech-communication, and psychology. It is particularly influential in interactional sociolinguistics, discourse analysis, and discursive psychology.

[Search for resources about conversation analysis](#)

What Is Grounded Theory?

The essence of [grounded theory](#) is that the researcher(s) begin, not with a pre-formulated hypothesis or articulated theory, but with the data generated by empirical social research. Instead of doing research to test hypotheses and theories, research is done to engage with reality as a way toward formulating hypotheses and theories.

Glaser and Strauss (1967) developed a method which is not purely empirical but requires a process of constant comparison of ideas and concepts with the actual material being generated in the research itself. The focus of attention is not on cases, in the form of individuals, institutions, etc. The researcher concentrates on incidents, for example interactions and events in the social world.

Strauss and Corbin (2015) have developed grounded theory in the direction of a more formal analytical strategy for engaging with qualitative data.

Glaser (1978) has rejected this approach, arguing that the crucial characteristic of grounded theory is the emergent character of theorization in an engagement with data. The term “[theoretical sampling](#)” describes the strategy of engagement where instances are sorted until nothing different is being discovered. The researcher achieves saturation in relation to the context and interests of a particular research project.

[Search for resources about grounded theory](#)

What Is Framework Analysis?

Framework analysis was developed in applied policy-centered research. Its underlying logic of inquiry is close to that of [grounded theory](#). However, reflecting its applied focus, there is:

- a more structured approach to the development of coding categories
- a more restricted conception of what are appropriate data for consideration in framing and developing theoretical accounts.

Framework analysis is problem-centered, rather than being concerned with wide and general theorization of the social world. It is appropriate as a qualitative research strategy when:

- The locales of the research are relatively similar
- Clear limits can be set on the time devoted to the research
- Clear limits can be set on the range of social action and processes which should be investigated.

[Search for resources about framework analysis](#)

What Is Visual Analysis?

[Visual analysis](#) is not distinguished by the form of the approach, but rather by the nature of the materials being analyzed. Simply put, it deals with images as the object of research.

Visual analysis in the social sciences is qualitative in form, and draws on approaches developed in art history. It is concerned with:

- The content of images
- The arrangement of elements within them
- The nature of the processes of production of the images
- The social context surrounding their production. There is a useful overlap here with [narrative analysis](#).

[Search for resources about visual analysis](#)

What Is Phenomenological Analysis?

The essence of [phenomenology](#) is “back to the things themselves.” This is coupled with an emphasis on the ways in which human beings in their life world understand and interpret the world for themselves. In general, any research approach which draws on the phenomenological tradition in philosophy and social theory can be described as phenomenological in character.

[Interpretive phenomenological analysis](#) is a qualitative research approach developed in psychology. It draws on:

- The phenomenological tradition’s concern with subjective human experience to explore cognition
- An ideographic approach to the development of research accounts of this experience.

[Search for resources about phenomenological analysis](#)

[Search for resources about interpretive phenomenological analysis](#)

What Is Content Analysis?

[Content analysis](#) is a process whereby any form of communication, including texts, images, etc., is subject to a process of counting based on:

- The frequency of certain words, phrases, and other linguistic sets
- The use of an established coding frame designed to generate measurements from qualitative materials.

Keywords or phrases can be used in searching large volumes of text in just the same way as they are used in bibliographic databases. This can lead to the identification of textual elements and also serves as a basis for counting frequency of occurrence.

More sophisticated content analyses pay attention to the ordering of elements in any communication or text.

[Search for resources about content analysis](#)

What Is Thematic Analysis?

Most methods of qualitative analysis employ a thematic approach in which textual, or other, sources are examined to identify themes. This can include field notes from ethnography, interview and focus group transcripts, and sets of images.

The usual stages of [thematic analysis](#) are:

- Preliminary scanning of the materials
- Development of a set of thematic categories, informed both by pre-existing understanding (often derived from the [literature review](#)) and emergent “in-vivo” inspirations
- Coding of elements in the materials as representatives of the themes
- Using a software package (e.g., NVivo) which enables the coding of textual materials to thematic nodes and sub-nodes.

The purpose of [coding](#) is both the organization of conceptual development and the indexing of materials for illustrative purposes in relation to presentation and argument.

[Search for resources about thematic analysis](#)

Read more about the [literature review](#)

[Search for resources about coding](#)

Which Qualitative Data Analysis Software Is Right for Me?

[This website](#) provides a detailed guide to choosing appropriate software to aid in qualitative data analysis.

How Do I Analyze and Interpret Mixed-Methods Data?

If you have used a variety of methods in your research, then you will need to do both of the following things in analyzing and interpreting your data:

- Analyze and interpret the materials according to the standard methods generated by the individual methods deployed.
- Integrate your materials in an overall analysis and interpretation directed toward answering your research questions.

You may have deployed particular methods in relation to particular research questions. If so, you can use the materials generated and analyze and interpret in a standard way for that method when answering each specific question.

However, even if you have done this, the research must stand as a whole and you will need to synthesize your findings in your overall conclusion. Sometimes you will have brought multiple methods to bear on a single research question or set of questions. This approach is generally described as *triangulation*.

The term “[triangulation](#)” is used in social research to describe the use of more than one method to generate multiple kinds of data about a particular research issue or set of research questions. It is a metaphor derived

from the land survey practice of establishing the location of a specific geographical point by reference to two other fixed points, thereby constructing a triangle since the two other points are fixed by reference to each other.

The English expression “what you tell me three times is true” conveys the purpose of this strategy. If different methods come up with the same account, then we have good reasons for believing that the account is accurate.

Here are three examples of triangulation in practice:

- **Exploring attitudes to a social issue:** You could use a quantitative survey with fixed questions and pre-coded answers AND qualitative strategies—for example, qualitative interviews and focus groups.
- **Looking at social change in a locality over time:** You could examine secondary data AND interview respondents in an oral history mode.
- **Exploring the effectiveness of a new pedagogical strategy for teaching basic statistics:** You could conduct a randomized controlled trial AND collect feedback information on experiences, possibly both quantitatively and qualitatively, from students and teachers.

[Search for resources about triangulation](#)

How Do I Analyze and Interpret Quantitative Data?

There are two basic strategies for engaging with [quantitative data](#):

- **Exploration**

Here you look at the data to see what they are telling you. You examine the data for patterns, signs of association, and examples of differences.

You do not approach the dataset with a pre-established hypothesis or model. But instead you work in a way which has much in common with the logic of [grounded theory](#) in qualitative work. (Strauss [1987] considered this approach was equally applicable to quantitative work.)

- **Explanation**

Here, strictly speaking, you have a hypothesis or set of hypotheses, which you develop in advance of any engagement with the particular dataset on which you are going to test them. You should not develop hypotheses on the basis of exploration of the dataset which will be used for explanation.

In practice, people don't often test hypotheses as models to see if they fit a dataset, even if preliminary exploration of the data set has suggested the outline of the model. Strict [hypothesis testing](#) is mostly done in experimental research.

Exploration and explanation are not mutually exclusive strategies. They are often deployed either simultaneously or in sequence in social research.

Your questions will require you to do any or all of the following things:

- **Describe**
When you describe, you say what things are like. We are often particularly interested in how things have changed over time. Describing change over time can be done using longitudinal data sets.
- **Test for differences among categories**
- **Classify**
When you classify, you establish categories based on data describing the cases.
- **Measure association/correlation**
When you measure association or [correlation](#), you identify the existence of and strength of associations among variates.
- **Regression analysis**
When carrying out [regression analysis](#), you model multiple relationships.
- **Factor analysis**
When carrying out [factor analysis](#), you identify underlying factors behind a set of data.
- **Estimate likelihoods**
When estimating likelihoods, you generally estimate parameters given incomplete information, e.g., data from a sample.
- **Make systematic comparisons across cases**
- **Identify connections**
When identifying connections, you establish the nature of networks.
- **Attach probabilities**
For sample data, you attach probabilities to the sample-based estimates of population parameters.

[Search for resources about quantitative data analysis](#)

[Tell me how to choose a statistical test](#)

How Do I Identify Variation and Difference?

[Variables](#) vary. We are always interested in how they vary from themselves. We are also particularly interested in how they vary among cases in relation to other attributes of those cases. For example, we might ask if there are differences in political voting intentions between men and women. “Between men and women” is the variable, and gender is the attribute.

We count the frequencies by value for cases and present this as a frequency table.

For ordinal variables we do the same. We do not consider the difference between the highest and lowest

value on an ordinal scale as the range since we cannot subtract on an ordinal scale.

We have:

- **Range:** The difference between the highest and lowest value. Obviously this is determined by extreme values.
- **Inter-quartile range:** The difference between the upper quartile (median of the top half of the range) and the lower quartile (median of the bottom half of the range). This is not affected by the extreme values.
- **The variance:** This is the sum of the squared differences between each value for all cases and the mean as the numerator in a fraction where the denominator is the number of cases. This is very much affected by extreme values.
- **The standard deviation:** This is the square root of the variance. Again this is sensitive to extreme values.

If we have all the cases and we see a difference in a variable in relation to the value of another variable for the cases, then the difference is there.

Or, if we have a sample and we observe a difference, then the difference may exist in the population from which the sample is drawn or it may be an artefact of the sampling process. So we have to test for the [significance](#) of the difference. To do this we have a whole set of statistical tests:

- **Chi-square test:** Tests for significance in the pattern of relationship between two nominal variables.
- **T-test:** Tests for significance of difference of means in small samples, with the necessary assumption that the continuous variable is normally distributed in the population from which the sample is drawn. The nominal variable is binomial. When the sample is large (greater than 40) the difference of means test can be used and the assumption of normal distribution of the continuous variable in the population can be relaxed.
- **Analysis of variance (ANOVA):** Tests for significance in the pattern of relationship between a multinomial nominal variable and a continuous variable. Kruskal Wallace's ANOVA does the same for the relationship between a multinomial nominal variable and an ordinal variable.
- **Analysis of covariance (ANCOVA):** This is like ANOVA but allows for more than one nominal variable as a control variable.

Analysis of variance and covariance are important tools in [randomized controlled trials](#).

[Search for resources about variation and difference](#)

How Do I Identify the Power of a Statistical Test?

The [power](#) of a statistical test is the probability that the use of the test will lead to the correct rejection of a null hypothesis when the alternative positive hypothesis is true. This is most important in relation to [statistical significance](#) when we are drawing inferences about populations from information derived from samples. In

formal terms, the power of a test is the probability of not making a [Type II error](#). That means falsely rejecting a true null hypothesis: the fallacy of affirming the consequent.

Power calculations are used to determine the sample size necessary to detect an effect of a given size in the population.

Note that we can never be certain that we have not made a Type II error, however big our sample size. In Karl Popper's (2002) terms, we can only fail to reject a hypothesis, we can never confirm it. So unless we have all the cases any statement made on the basis of sample data about a population may still be wrong. We just attach a probability to the likelihood of this happening.

[Search for resources about statistical power](#)

How Do I Test for Significance?

If we make statements about a population on the basis of information drawn from a sample of that population, we can never be certain that the statements are correct. The value of a variable as measured from a sample (a statistic) is not necessarily the value of that variable in the population (a parameter). Differences in the values of a variable in relation to other attributes of the cases which we observe from the sample data may or may not exist in the population.

We cannot assert that our sample tells us about the population. However, if we have drawn a proper random sample of a sufficient size from the population, we can attach a probability level to our statements about that population. This is the very useful consequence of the law of large numbers. It allows us to infer from samples to populations.

All statistical tests attach significance levels to the statements they generate about population parameters/differences on the basis of sample data. These can take the form of:

- **The confidence level:** Estimates of the range of a value for a continuous variable.
- **The significance level:** Estimates of the likelihood that a difference or association observed in the sample actually exists in the population. Significance levels are usually given at the 5% or 1% level. This means for 5%, if we drew all possible samples of size N from the population, then no more than 5% of these samples would give this estimate of value, difference, or association if the value was outside the range, the difference did not exist, or there was no association.

[Search for resources about significance](#)

How Do I Identify Key Factors?

We often have a very large number of measures for our cases and may have reason to think that the observed measures are actually indicators of a smaller number of causal factors which lie behind these observations.

For example, we might test mental capacity using a variety of different testing procedures which generate scores. We might think it possible that there are one or more causal factors lying behind these scores but that

the number of such factors is fewer than the number of scores we have. We can identify the key factors in two ways.

- **Principal components analysis:** This is an exploratory procedure which takes a set of correlated variables and generates a smaller set of uncorrelated components by mathematical transformation. It is a useful data reduction technique. It makes no assumptions about the possible existence or nature of underlying components.
- **Factor analysis:** Here there is a pre-existing assumption that observed variation among a set of variables is actually the product of underlying variation in a smaller set of causal factors—the **latent variables**. The approach uses regression techniques to identify these latent variables and generates error terms. It is a confirmatory/hypothesis testing method.

[Search for resources about principal component analysis](#)

[Search for resources about factor analysis](#)

How Can I Use Descriptive Statistics?

Descriptive statistics are ways of summarizing variables in a useful way. They deal in:

- **Frequencies:** Counts of the number of cases which have a value. These can be used for nominal, ordinal, and continuous variables.
- **Levels:** Descriptions of the middle of the data range. Averages.
- **The median:** The value of the case which has an equal number of cases with larger and smaller values. It can be used for ordinal and continuous variables but is most useful for continuous variables.
- **The mean:** The middle of the data in terms of the values of the variable for the cases. Add up the values of all the cases and divide by the number of cases. It can only be used for continuous variables and is sensitive to extreme values.
- **Spread:** Measures of spread describe how similar or varied the set of observed values are for a particular variable. The most common measure of variation, or spread, is the standard deviation. Other measures of spread include the range, quartiles and the inter-quartile range, and variance.

[Search for resources about descriptive statistics](#)

How Do I Identify Clusters?

When we have a large number of cases and lots of quantitative information about the cases it is often very useful to be able to sort the cases into categories based on that information. The two ways in which we can do this are:

- [Data mining](#)
- [Cluster analysis](#)

Here we use data mining tools, usually based on neural nets, to search very large data sets—[big data](#)—and generate categories. These tools are based on training the neural nets with learning algorithms.

This term describes a set of procedures based on matrix algebra approaches which enable us to sort cases into kinds on the basis of their similarity or dis-similarity. There are a number of different algorithms which can be used for this and most are available as options in standard statistical packages. It is possible to sort on the basis of continuous data, categorical data (ordinal data are usually regarded as categorical here), or a mixture of both.

Cluster analysis approaches can be:

- **Hierarchical**

Start by putting together the two most similar cases, and continue step by step until all cases are in one cluster. This style of approach generates error messages, which indicate when very different things are being put together. It is useful for identifying levels of useful typologies. We can see what size of sets of types really looks different. Are there three, seven, or nine meaningful categories in the set of cases?

- **Agglomerative**

This is typified by SPSS Two Steps clustering. Here there is no hierarchical fusion. Instead there is a single pass at the data which sorts them into the number of cases which best generate a differentiated typology.

- **K means**

Here the number of clusters is known in advance so this approach does not generate a new typology. Rather, it is a way of sorting a set of cases into an existing typology and describing the character of the categories generated.

All clustering approaches generate descriptions of the clusters in terms of mean values and standard deviations for continuous variables, and frequency counts for nominal variables. Therefore, it is possible to write a textual description of each category generated.

[Search for resources about data mining](#)

[Search for resources about cluster analysis](#)

How Do I Estimate Likelihood?

The technique of [maximum likelihood estimation](#) is used to estimate the value of a parameter given data from a sample.

The term likelihood is not synonymous with probability when used in statistical inference.

Probability describes the possibility of something, given existing knowledge. So in statistical inference we rely on our knowledge of sampling potential, given the law of large numbers, to attach a probability statement of statistical significance to a statement based on sample data.

Likelihood describes the potential values of the population parameter, given the relevant sample value.

[Search for resources about maximum likelihood estimation](#)

What Are Model Processes?

All **models** are representations of some aspect of reality which are simpler than the thing being represented. To be useful, models must incorporate all the essential features of the thing being modeled.

Process is a term used to define a sequence of actions toward an end.

In business, there is a well-developed set of techniques which work by analyzing the elements of an organization's actions toward an objective and representing them in quantitative form. We can model in **simulations** which are a special way of attempting to represent processes as they develop forward through time.

The approach of *process tracing* was developed in political science as a way of retroductively exploring how social systems came to be as they are in the present. To do this we use systematic comparison coupled with an element of quantitative modelling.

[Search for resources about process tracing](#)

How Do I Create Simulations?

Transcript

00:15

Thank you very much for agreeing to talk to us today. What I wanted to ask you was what, methodologically speaking, is simulation? Well, simulation is a new way of thinking about things.

00:35

It's not really a method in the same way as linear regression is a method. What simulation is at its heart is the idea of taking some ideas about society and building, in a computer, a model, a simulation, of that,

01:01

of the overall ideas, so that you have essentially a virtual society. The great advantage of that was really two great advantages. First of all, if you managed to it, then you know that your theory or ideas

have some sort of degree of coherence and completeness

01:22

because otherwise you wouldn't have been to do it in the first place. And secondly, it has the advantage that you can then do, what I call, virtual experiments. So Instead of doing experiments on real people, which is obviously expensive, unethical, all sorts of problems, you can do experiments

01:43

on virtual people. And that can be both very instructive and, I must admit, also quite fun. What do you mean by a degree of completeness? How would one know that you have a degree of completeness? Well, actually, simulation models

02:04

vary a great, as you might expect, from the ones that are really abstract and very simple to ones that are highly complicated and try and be as, what I call, facsimile models. Like a fax, they're trying to be as exact a copy of the real thing as is possible.

02:27

Both of these sorts of simulations have their advantages. If you have a nice, simple field, you've got perhaps a theory about social mobility, that stratification kind of happens through a process

02:48

of education and occupation. You can express that, as people have indeed done, in relatively straightforward terms. And you can build a simulation, and you can see whether that theory kind of works out in your simulation. Do you get social classes, for example,

03:09

coming out of your model? Now, when I say it's a very simple thing, we did not expecting actually that the number of plumbers in our model will correspond exactly to the number of plumbers in reality. What we expect to see is social classes.

03:32

And if we find those, then that's good, and it gives us a little bit of confidence that our original theory is doing the sort of things that we thought it was going to do. At the other end of the scale, if you like, imagine somebody who is being hired by a city council to tell the city council about the phasing at traffic lights--

03:58

how long should the red light go on for, and how long should the green? That's a classic example, actually, of a simulation, where we're simulating cars going down streets, and we're building to that model lots and lots of information about the geography of the city,

04:20

about where people live, where they all work, where they shop, blah, blah, blah-- lots of data. And we need all of that data because otherwise when we come to use the model to test kind of what is, what happens if we make the red lights go on for longer, will we get less traffic jams.

04:43

We can't even really answer that very specific sort of question if we've got all this data, if we've got our facsimile model. So [INAUDIBLE] the great thing about social simulation is that there are all these different opportunities for ranging from the [INAUDIBLE], very theoretical, very conceptual sorts of models, to the very practical,

05:06

very, very applied. So what social processes can and cannot be simulated? Let me take your question apart a bit because I think I want to emphasize a couple of words there in your question. Social simulation is good at modeling social processes.

05:33

And that's actually a bit of a change from most ways of modeling, most ways indeed, most research methods in the social sciences, when you come to think of it. What do I mean by social process? I mean something which has got an intrinsic time dimension to it.

05:55

So, for example-- we'll go back to my earlier example. We might have a model about social mobility. And with that, we would model the processes of getting an education, of making decisions about jobs, about being in jobs, about retiring and so

06:17

on over the life course of lots of different people. And what we do in most simulations then is to model at the individual level what individual people do-- individual agents, as we call them.

06:38

And then we look to see what the global, macro societal level of the outcomes are. But, to come back to your question, contrast that with, let's say, the kind of classic work on social mobility,

06:59

which takes a table of fathers' and sons' occupations. Although there is a sense of time in that-- it's fathers and sons-- we aren't really putting into the analysis the processes by which fathers' occupations, if you like, become the sons' occupations-- not in the same way, at any rate.

07:20

And that's really the exciting thing about this sort of social simulation is that we can theorize about and think about social processes, and then build them into our models. Your question, to go back to the question, was what kind of social processes can we

07:40

or can we not simulate? And the answer to that is rather an ambitious one, I suppose. I don't think there are any social processes we can't simulate. And indeed, if we look at the literature now in social simulation, it's extremely varied. And the sorts of things that people simulate

08:03

vary from migration, segregation, industrial development and innovation, education. Well, it's just an endless list of things. It's with a bit of imagination, we can do lots of things.

08:27

But can you simulate the life course of an individual? Can you simulate the other end, both on the micro and the very macro end of very complex social processes, where there are an awful lot of things going on, an awful lot of variables? Well, the essence of modeling, in general, indeed you might say, the essence of most social science is to simplify.

08:52

If we wanted to model the real world, the social or physical world in exact detail, well, we'd have a model that was as complicated and as difficult to understand as the real world basically. That's a pretty pointless exercise. So inevitably, even in some of the most detailed

09:16

ethnographic kinds of methods, you have to do some kind of simplification. There's some things that you can't observe, and some things that you're not interested in observing. So we're always inevitably engaged in some sort of simplification and abstraction. And the same is true [INAUDIBLE] simulation.

09:39

What we have to do if we are modeling social processes is we have to say, well, what kind of specific questions are we actually interested in so to define the scope of the model. And then secondly, what are the essential aspects of those social processes that we're interested in modeling?

10:01

Let's try an be as simple as possible. There's that thing about, what's known as Occam's razor, which says that a simple explanation is preferable to a complex explanation, that the simple one provides as good of answers, as it were. Occam's razor is a very useful razor

10:22

in any kind of social science. We want to be as simple as we can. So if I'm a new researcher, and I'm interested in social simulation, what sort of things do I need to bear in mind in posing my research question and thinking about my research design?

10:43

Well, actually, although social simulations sounds like a really perhaps, well, new, innovative, and perhaps even weird way of going about things, but the way and the research process is very similar to other kinds of method.

11:03

It starts off with the crucial and often the hardest part is to define clearly what the research question actually is. And that's common to any method. Once one has done that and looked at what the literature has already said on a matter,

11:25

then it is a matter of thinking through, theorizing, the processes that you're interested in and gathering together some appropriate sorts of data. And social simulation is essentially an analytical method. It's not a data collection. I mean, you still need survey data or ethnographic data

11:48

or documentary data or whatever can be collected in conventional sorts of ways. So the next step, and the one that perhaps is a little bit different from other methods is that we need to put together a model. And that used to be something that only computer geeks could

12:09

manage. But nowadays, there is now software that will run on your PC which is really relatively easy to use. And we run courses in which people have never done any computer programming or anything like that before and within a day or two, writing

12:35

their own computational models. So it really is not that hard anymore and getting easier by the day. Once you've got your model, then there's a process which we call validation, which is essentially comparing the outputs

12:57

of the model with the data that you've observed in the real world. And validation really gives you the confidence that your model is the right model. And you may have to go through this process iteratively because it probably may start out [INAUDIBLE]. It won't be right, but you can compare,

13:20

change your model, and so on. Once you are happy with your model, the you'll probably want to do what I refer to as experiments, or you might call them what-if exercises. So what would happen if there was positive discrimination?

13:42

It all depends of course, on the domain, but the kind of conclusions that you would want to draw from your model are probably [INAUDIBLE] going to come out of that kind of trying out things with your model now that you've got it. And sometimes those are conceptual theoretical conclusions, and sometime they're quite practical,

14:04

positive conclusions. It was interesting for me that you mentioned ethnographic work and that that can be modeled as well. If I'm an ethnographer, do I have to somehow turn my data, which is qualitative data, of course, into some kind of numerical data to do this work? I should have made more of that point-- thank you

14:25

for waiting-- because actually, us social scientists tend to divide methods into the quantitative and the qualitative. And simulation doesn't fit into that category, or you might say it's in the middle between because it's perfectly capable of dealing with both quantitative and

14:46

qualitative data. So we might, if we were a qualitative researcher, have some kind of ideas about people in this situation in this culture do things like that. And we can take a statement like that

15:06

and build it into our model. And our agents and our virtual people, we can make them do things like that when they find themselves in this particular situation. But we also have the quantitative researchers, or survey researchers [INAUDIBLE] I know that 43% of respondents vote Labour are aged 52 or whatever,

15:33

some kind of quantitative data like that. We can build that sort of thing into our model too, so we can make our agents vote one particular way to such-and-such percentage and have such-and-such an age distribution. So that's really helpful and useful to have this sort of thing. We bring together the qualitative

15:54

and the quantitative. At the same time, in away, the whole thing about simulation is that it does have this kind of integrative purpose that it can bring together different disciplines as well.

16:15

I think it's probably-- even without going into detail, it's quite easy to see how, for instance, you could build a model that incorporates some ideas from sociology about, let's say, social networks and relationships and friends and so on that could build in some ideas from economics about how people price things,

16:37

about resources, and so on and so forth. And you can straightforwardly build models where the agents have some sort of notion of geography, that put him on a map, if you like-- that is spatiality, as it's called, in your model-- and that the agents have

17:01

some lots of notions of beliefs and culture and so on. So we can put in some anthropology and even some psychology. So actually, if you look at work that is going on in social simulation, it's very often the case that that

17:22

is being done in an interdisciplinary fashion, that the model serves as the talking point for fellow

researchers from a number of different disciplines who are collaborating together.

A [simulation](#) is a special form of model which works in the same way as the system it is supposed to represent. Though simulations have been made using mechanical means in the past, almost all simulations now are constructed in digital form using computer technology and software.

Simulations in social science fall into two categories:

- **Equation-based**

Here social processes are represented by terms in a mathematical model which comprises a set of equations. These equations are considered to describe the set of causal processes in reality. As with all mathematical models, the essential elements are the functions in the form of the equations and a set of initial values for the terms in the equations. Equation-based models can incorporate elements of structure as well as take account of the behavior of individual social actors.

- **Agent-based**

[Agent-based simulations](#) are driven not by equations but by sets of rules, which are observed by entities in interaction with each other. The entities behave according to these rules and the nature of the agents with which they interact. More sophisticated agent-based models can allow for some learning by the agents on the basis of the experience of previous interactions. The underlying methodological premise of agent-based approaches is micro-emergent. Things emerge from the interaction of agents and the whole of social structure is post hoc emergent.

Simulations are important tools for saying what will happen if things are like this, i.e., for predicting.

A distinction must be drawn between simulation models which are purely deductive and those which are in some way calibrated against empirical reality through the use of data describing that reality.

- **Purely deductive models** have no way of demonstrating isomorphism with reality, correspondence with reality, other than through the assertions made by the model builder. They are in effect fictions.
- **Models which are calibrated** do have a connection with reality which depends on measurement of existing reality.

[Search for resources about simulations](#)

How Do I Identify Change?

Change can be incremental or radical. It can be a change of degree or a change of kind. Both are of interest to social scientists but we are particularly interested in changes of kind.

We generally think of changes as occurring through time so we find it useful to use [time series](#) of data. These

series might be:

- as simple as a before and after intervention pair of measurements across cases in an experiment in relation to a single variate outcome value
- more complex involving repeated measurements through time across cases for a very large set of variables.

We can identify changes in the value of single variables by inspection. It is often useful to visualize these by preparing a time ordered graph.

When we have very large numbers of variables it is often useful to conduct time ordered [cluster analyses](#). This enables us to identify when cases change the kind of thing they are, and also changes in the very typologies, sets of categories, themselves.

There are a range of *time series* methods, mostly derived from Econometrics, which can be used to model changes through time going backwards. These can be developed as predictive tools. However, when there is [nonlinearity](#) in the system, prediction becomes much more difficult.

Transcript

00:00

[MUSIC PLAYING] Rachel Thomson and Julie McLeod Thank you very much for talking to me today.

00:22

What I wanted to ask you is how would I go about researching social change? Well, social science, as a discipline, is very concerned with social change and theorizing social change. But in terms of actually researching it, we don't necessarily have great methods for it. In fact, the discipline history has to be more well developed

00:44

for that. In the quantitative tradition, the best way, really, to search social change is being found in the panel studies, the longitude and the panel studies, which have been a really important, almost natural experiment, really, of watching people, cohorts of people over time and see what happens to them the way I've related to historical and biographical

01:06

change. In the qualitative tradition, there are a number of threads, really, of researching social change, which themselves come from different bits of sociology, cultural studies, and history. And one way that we can think about them is by considering them in terms of methods which

01:26

are about memory and remembering, methods which are about really the unfolding of time, and working in the present tense, and methods which were about ideas of inheritance and change in the future. And these different approaches can be done on different scales. So the biggest scale, of course, is the historical scale.

01:48

We're talking about social change. The smallest scale might be the unfolding of the day of a lesson in a school. So we can think about social change in terms of processes of many different methods. I think the other dimension of it is also the relationship between the biographical and the broader sort of cultural sweep.

02:09

And some of the methods that are useful for that are ones which try and make the bridge between the experience of changing counter-biographically and the way in which that might be a record to understand larger social processes, and larger movements and social change. So some of the methods that are useful for that include oral and live history, where the individual story

02:30

is told as a way to understanding that person biographically and generational [INAUDIBLE], but also as a window onto understanding all the social processes. Equally, memory work, where memories are used as a prompt, as a trigger into opening up reflection on the past.

02:51

But also, in both methods, there are ways of understanding the past, but also a window onto present concerns. And in both, it's that movement between the larger social and the biographical, it's an issue. Yeah. Can you tell me more about memory work? How would I do it? What specifically does it involve?

03:12

Well, there's two versions of memory work, one with a hyphen, memory-work, and one without a hyphen. The one without the hyphen's much more open. And really, you can use memory in all sorts of ways. But there's a specific technique called memory-work which was used initially by Frigga Haug and colleagues in the early '80s, and was developed

03:33

by a whole range of academics, including June Crawford and Kippax in Australia in the '90s. And we

looked in detail at how that works. The basic process of memory work is coming up with a trigger word. It could be anything. It could be child. It could be mother.

03:53

It could be home. It could be bacon-- any word. And using that trigger, you write a memory in detail, but not in the first person. So you don't say I. You say he or she. And then these texts become the basis for analysis. So if we're going to get together, all having produced some memories in the same trigger.

04:13

And then they analyze those memories without going into autobiography, but to try and attempt to understand from these texts what we can know about the time and place that they relate to. So for instance, childhood has been used a lot in childhood research. But also, what it can tell us about the sorts of investments that adults have in a particularly struggled period

04:34

in their childhoods. Where does that come from? Does that come from a kind of psychotherapy kind of trajectory? Well, its roots are really in the consciousness phrasing methods of feminism in the '80s, and the '70s and '80s. That's where it began. But there are a number of threads, and it's been employed in different ways. So for instance, in Annette Kuhn's memory work.

04:55

But she works primarily with film, photographs on her own, not with a group. It's influenced very much by psychoanalysis and within cultural studies traditions. So there isn't one memory work, but there are a number of disciplines for whom memory is a very important resource.

05:16

So when you think of oral history and life history too, we can locate the origins of those sorts of methods in social movements such as feminism or movements which are trying to listen to what people say, gain insight into people's experiences, trying to tell another story, a counter story to the official historical record.

05:37

So oral histories-- again, there's no single way of doing it. But one of the key features of it is an extended interview with someone. Sometimes the emphasis is more on understanding the particularity of that experience to try and delve deeper. It has sometimes in its elements, a sort of memory work

05:58

process. Other emphases are when the focus is much more on the experience someone might have had growing up through the war or to try and tap into the experience of a particular era. One important aspect of oral history

06:20

is its potential to shed light on political issues in the present. In our book, *Researching Social Change*, in the chapter on oral and life history, we take the example of oral history as a form of testimony, and we discuss the experience of aboriginal people in Australia providing testimonies or experiences

06:41

to a commission investigating what's called the Stolen Generation. From the 1930s through to the 1970s, many aboriginal children could be forcibly removed from their homes and put with white families or in orphanages. And part of the inquiry was to get the experiences and the effects of this intervention

07:03

on the lives of aboriginal people. So therefore, oral history becomes a very, very important way of recording that history and putting on the official record the perspectives of people who traditionally have been silenced. Why couldn't you use other methods to do that? Why would you choose oral history to do that? That's a very good question, and one

07:24

of the advantages of oral history for this group of people was that so much had been written about them by other people that there was a reluctance for them to be researched by others, to have, for example, traditionally white researchers or white anthropologists come in and tell their story. So the telling of their story had a political dimension

07:48

to it, as well as a very deliberate way of getting their words. So it was a way of trying to catch a change, the experience of the change, biographical change, the experience and cultural change, the experience of the impact on aboriginal communities through the vivid and direct voice of the people who were most affected by that.

08:09

So it's true. It could be complimented by other types of sources, but oral and life history give an immediacy to the experience of change. How would you distinguish between oral history and life history? Do you want to say something? Well, I don't think there are hard and fast distinctions. More than anything else, it's probably disciplinary.

08:30

Some people with the historical training would think and describe themselves in a tradition of oral history. Life history is more sociological tradition, and they are very different kinds of sets of ethics and practices. So interestingly, within live history, there's perhaps still, in keeping with sociology,

08:53

concerns around confidentiality, concerns about whether you can archive this material, or the oral history tradition. People have always seen it as a testimony. Do it in their own names. And the material is created to be archived. Yes. So there's a sense in which it's not necessarily even created to be analyzed. It's created for the record.

09:14

Yes. But in practice, sometimes they might look very similar to each other. Yes. I think that's right. And often, some of the benefit of the oral history is seen as a record, a descriptive account of something rather than necessarily producing a lot of interest in closed textual analysis of the hidden meanings, of the latent meanings of what's said,

09:35

but merely a lot of attention to the [INAUDIBLE] story. But in the last several decades, there has been, in both oral and life history, interest in the relationship between the interviewer and the interviewee. So that then becomes an important part of that research method. It's not simply a matter of the researcher

09:56

going in and getting the story, but the dynamic between the participant and the researcher, which also is part of the story. So if I was your PhD student, and I wanted to use these methods myself, what would be your do's and don't's, in terms of advice for me going out and using these methods?

10:18

Well, the methods themselves always have batteries, take the quota. We've all undertaken the most extraordinary interviews and found that they have been recorded, so I would say probably more. The most important thing is being technically right. And then interestingly, in the oral history tradition, a huge amount of emphasis is put on that, including--

10:41

the norm is to where if I was recording somebody, to listen, to being attending to quality the whole time with the view that this is a record, an important record, and the quality is really important. So yes. That's

number one. Of course, the key really is accessing in your group

11:01

and who they are, getting that right, finding people whose stories are important, whose stories need to be told, and being very clear with them on what grounds you're doing this, and being clear with yourself, actually, probably is always the most important thing, and then you can communicate that to others. I would say an important part of the method

11:24

is to think that the interview is not-- the oral history and research method is not simply the doing of that interview, and that the important thing is to allow time immediately after the interview to write your initial field notes as it were. So you've got that record, but that's not all that is happening. So it's making time to write down your impressions,

11:47

to then compare them, if you can, to transcribe the interview yourself or to listen to it with a digital recorder, to listen, and to immerse yourself in that interview, and go back to those initial notes. Often, those initial reactions you have will have the germs of subsequent analysis, and I really cannot emphasize enough to allow that reflective time immediately after

12:10

the interview. OK. We're coming close to the end of the time we have. But I wondered whether you have anything else that you would like to say about how we research social change or whether you're happy with what we've said now. Well, I think it'd be good to maybe say a little bit, even very quickly about the other part, still the sort of time

12:33

registers that you can work in, because we've talked about memory and working with the past. But also in our book, *Research And Social Change*, we also talk about the idea of working in the present tense and capturing the present tense as it happens and unfolds. And for instance, the methods like qualitative longitude and research, where you re-interview people,

12:55

or you might use observations. You keep going back over time, looking at how things change, how people change, how people contradict themselves. It's very interesting, making sense of that, but also methods like ethnography, where you the duration is your time in the field. You're immersed in a situation that is changing and unfolding.

13:16

And those methods-- the particular quality of data that they give rise to, the extent to which the researcher is absolutely implicated in that data. You can't have an external search. The research is part of that, and it has a particular flavor and taste, the kind of research that's done in that mode. But we also talk about methods of research

13:38

which are more future oriented, even though they involve going back. So one example of that is revisiting studies, so where researchers-- it may be the same researcher at the beginning of their career, at the end of their career. They go back to a study that they did before, and they may contact the same search subjects. They might go back to the same institution

13:58

and find out both how they've changed, how the place has changed, and how the framework for understanding has changed over time. So that's a very interesting example. And then-- sorry. I just have to say with that, and also with the example, the method of qualitative longitude and research, for some people beginning a doctoral study,

14:21

it may seem a difficult method, because they might presume a long period of time is needed. And the important thing with revisiting studies and with qualitative, longitudinal studies, is that you set up that you plot to observations or interviews at different stages. It doesn't necessarily have to be over a long period of time.

14:42

It's how you frame the question and the return to the interview that's important. So you can do a study which has a longitudinal method and a longitude of dimension in a relatively short period amount of time. And increasingly, PhD students are using longitudinal method. So would that be to say, secondary data or archival work, as well?

15:03

No. They're original research. So for example, you might want to look at something-- three points in a process. The process might be a schooling process. Yep. It might be an identity change process. It might be before and after the birth of a child. It's a very good method for looking

15:24

at change and transition. The transition points and standard, critical moments. So the shift from the end of secondary school to university or working life. You might be able to do that over an 18 month period. If you can do that over a three month period, it's rich,

15:45

but if you identify that points that you're looking at as change, change points, that's the challenge of the method. But there are some ways of researching social change which you can do right here and now. For example, one of the methods, perhaps the last one we talk about is intergenerational research, where you could capture historical change

16:07

by looking at the here and now at members of different generations. So this can be done, quite obviously, in families. For example, I've been involved in a study where we looked at three generations of mothers. So new mothers, when they just had babies, their mother and their grandmother. And you're looking at them all in the same moment.

16:27

So they're all reacting to the birth of one child, but reacting in relation to their own experiences, which may have been in the '50s. It might have been in the '80s. And that's a way of capturing in a particular moment in time both change, but also how the past is always present, particularly in the politics of the present.

16:48

So some of the politics of motherhood now are shaped by the memories of those who did it in the '50s and felt there was plenty of time, and none of these gadgets to make it all complicated. So there's lots of ways of getting at change, because change is part of everyday life. And in a sense, you can always find a routine.

17:10

And what we've tried to do is show some of the rich qualitative traditions, that people can draw on to access that. [MUSIC PLAYING]

[Search for resources about time series analysis](#)

How Do I Identify Networks?

We often can construct data sets which describe connections among cases of interest to us. *Big data* derived from metadata sets describing communication connections through electronic means can generate enormous accounts of connectivity of this kind.

We can use [social network analysis](#) which draws on the mathematical repertoire of graph theory to describe such networks overall and in terms of particular sets of connections within them.

[Search for resources about networks](#)

[Search for resources about social networks](#)

How Do I Make Comparisons?

The [comparative method](#) is one of the most powerful tools of the social sciences in searching for cause(s). The essential logic was described by John Stuart Mill (1843) through his elaboration of the method of similarities and the method of differences. It is the latter which is the basis, albeit much modified and developed, of systematic comparison methods in contemporary social science.

The most important of these is [qualitative comparative analysis \(QCA\)](#) which uses a set theoretical approach and software based on the application of De Morgan's Law to identify multiple configurations of complex causal sets in relation to outcomes. QCA has crisp set and fuzzy set variants. It allows for both multiple causes (equifinality) and complex interactive cause.

[Search for resources about comparison](#)

How Do I Test for Causality?

There is no such thing as a test for [causality](#). We can only observe associations and/or construct models which may or may not be compatible with what the data set(s) tell us. Remember always **CORRELATION IS NOT CAUSATION**. If we have associations in our data, then there may be causal relationships between variables but there may be some intermediate cause or there may be no real causal relationship at all.

A good example of spurious association is the high degree of correlation between the number of Australians attacked by sharks on a given day and the tonnage of ice-cream eaten by Australians on that day. We might infer that eating ice-cream causes shark attacks; given the likely sequence and the infrequency and severity of shark attacks, the reverse is not plausible. However, it is obvious that there is a common cause—hot weather. People eat ice-cream when it is hot, go to the beach, go into the sea, and are exposed to shark attacks, which are more common in hotter water.

If there is no association, then we can generally rule out any causal relationship. If there is association, we cannot assume a causal relationship.

[Search for resources about causality](#)

[Search for resources about correlation](#)

How Do I Identify Association Between Variables?

We are very interested in associations among variables. By this we mean that variables vary together in some systematic way. Generally, we call this association, and for continuous data we call it [correlation](#). When there

is association **there might be** a causal relationship, but remember always that **correlation is not cause**. The association might be spurious with no causal connection whatsoever.

Statistical tests of association include:

- **Contingency coefficient and Cramer's V** for associations among nominal variables.
- **Spearman's rank order correlation coefficient** for associations between ordinal variables.
- **Pearson's correlation coefficient** for associations among continuous variables.

Typically, measures of association are constructed so that they have a value of 0 when there is no association, 1 when there is a perfect positive association and -1 when there is a perfect negative association. That is, as one variable increases the other decreases.

For Pearson's correlations, the measure of the strength of association between variables is the square of the correlation coefficient.

It is possible to construct:

- Multiple correlation coefficients where there is a set of more than one "independent variable" and one dependent variable
- Partial correlation coefficients which measure the association between two variables after having "controlled for" their common variation with one or more other variables.

[Search for resources about association](#)

What Is Regression Analysis?

In statistics, [regression analysis](#) is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables. However, this can lead to illusions or false relationships, so caution is advisable. Remember, correlation does not imply causation.

[Search for resources about regression analysis](#)

What Are the Different Levels of Measurement?

It is crucial to establish the levels of measurement of any sets of quantitative data you have. This determines the kinds of statistical and related operations you can carry out using that data.

Remember, it is very likely that your data set will be "mixed" in terms of levels of measurement and will contain

all of nominal, ordinal, and continuous data. Important statistical techniques are concerned precisely with establishing relationships among data at these different levels. In particular, they deal with relating categorical/nominal data to variation in ordinal and continuous data.

- **Nominal (also known as categorical):** Here the number is just used to stand for the name of a category.
- **Ordinal:** Here we can rank and use the numbers to indicated ranks. BUT we cannot carry out any normal arithmetic operations because ranks have no quantitative content. A rank of four is not twice a rank of two.
- **Continuous (also known as scalar):** Here the number stands for a quantity which can be manipulated by all the normal arithmetic procedures of addition, subtraction, multiplication, and division.

Numbers which look continuous may not be describing a truly continuous variable as that variable is expressed in social life. Age, income, and wealth are variables which appear in a continuous form but might be better thought of in terms of ordered categories.

For example, someone aged 30 is three times as old as someone aged 10. But the ages really describe stages in life. We generally deal not with “raw” age as a variable but with an age category which relates to those stages, such as 0–1 infant, 2–5 young child, 6–12 older child, 13–16 adolescent, and so on.

[Search for resources about variables](#)

How Can I Interpret My Data?

All data have to be interpreted. We have to think about the data and decide what they mean. This is just as true for qualitative data as it is for quantitative data. We often talk about [quantitative analysis](#) which means the use of mathematical techniques and particularly statistical techniques to generate results. We also have to take account of computer packages which generate pictures and so present quantitative information as images, in the form of graphs and related output. This is just as important as actual numbers themselves.

But both the images and the results have to be interpreted. We have to say what they mean. Numbers do not simply speak for themselves. So the tools for the interpretation of data include:

- Processing through a computer-based package to generate numbers and images from quantitative data
- Reading through and organizing qualitative data, often by a process of thematic coding. We can use computer-based packages (e.g., NVivo) to do this. Major frames of reference for this process include [grounded theory](#) and [analytic induction](#).

[Search for resources about quantitative data analysis](#)

How Can I Eyeball Statistical Data?

Eyeballing statistics is the process of looking at a set of data and making estimates of statistical values without carrying out statistical calculations. It can be done by looking at tables of “raw” data. It is much easier to do if you have the data presented in graphical form. Almost all statistical packages now contain elements which allow you to construct graphs of any form.

To say something useful about single variables, you can use:

- Histograms
- Pie charts
- Frequency graphs.

For multiple continuous variables, you can see a lot in comparative terms by constructing a set of side-by-side box plots.

[Search for resources about eyeballing statistics](#)

What Are My Conclusions? Do They Respond to My Research Questions?

Your research questions should structure the organization of your working through your research materials toward your conclusions.

Remember, it is perfectly appropriate for new questions to emerge at any stage in a research project prior to the writing up of your conclusions. This may happen when you are doing the research and/or when you are working through the materials generated by the research. You can go back to the point where you specified the original questions and add the new ones, so long as you make it absolutely clear how and when you came to formulate them.

The best way to conclude a piece written on the basis of a social research project is by answering your questions:

- Individually
- In synthesis as a whole.

This does not mean that the conclusion should be written as question one, followed by answer one (although that can work quite often). Rather it is a matter of keeping your questions in mind and structuring your conclusion with reference to them.

Read more about [writing up my research](#)

[Tell me how to choose a statistical test](#)

[Use this checklist to decide which software to use to carry out your data analysis.](#)

Checklist: Essential First Steps for Analysis

Here are some of the questions you need to ask yourself before beginning your data analysis, and some

things you need to take into account.

- **Is this an appropriate point to begin analyzing my data?**

It can be appropriate to begin analysis before completing collection of all the data. But you have to remember that new findings may require a change in those analyses developed before they are available.

- **Have I worked out my overall analytical strategy?**

You may very well use more than one analytical/interpretive method in dealing with your data. But you need to have an understanding of how these will work together in order to answer your research questions and draw appropriate conclusions from your research.

- **Do I have a working command of the framework(s) of understanding and method(s) I am going to use in analyzing my data?**

Your methodological position will have a profound influence on the ways in which you analyze and interpret your data. Methods by which you construct data are largely epistemologically neutral. This is not the case for analysis/interpretation. You need to be clear on this. You also need to have a decent command of the actual technical aspects of the data analytical/interpretive approach(s) you intend to employ.

- **Have I got the resources I need to carry out my analyses?**

The primary resources here will take the form of access to appropriate computer software which can handle the materials you have in an appropriate way. You will need a statistical analysis package and/or a qualitative data analysis/management package. These packages differ in terms of the range of tools available within them and, for qualitative analysis/interpretation packages, of the underlying framework of explanation which is built into them. You need to think carefully about what you want to do and how you want to do it and then make the right software choices.

Checklist: What Software Should I Use?

Here are some questions to ask yourself when deciding which software package is most suitable for analyzing your data?

If your data are quantitative:

- **Do you want to carry out operations based around variables?**

If yes, do you prefer to use a graphical user interface?

IF yes, then SPSS will work for you.

IF no, then consider using R or Stata.

- **Do you want to work in a set theoretical frame?**

IF yes, the Fuzzy Set QCA software will work for you.

If your data are qualitative:

- **Do you want to work by establishing thematic coding working in a way generally congruent with grounded theory?**

IF yes, the NVivo will work for you.

IF no, then ATLAS.ti offers greater flexibility in the use of textual elements.

See [this link](#) for further assistance in choosing suitable software.

[Tell me how to choose a statistical test](#)