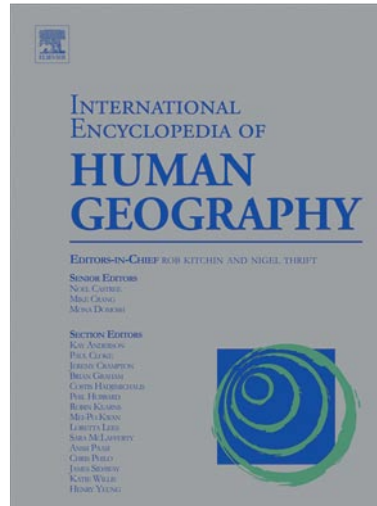


**Provided for non-commercial research and educational use only.  
Not for reproduction, distribution or commercial use.**

This article was originally published in *International Encyclopedia of Human Geography*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at: <http://www.elsevier.com/locate/permissionusematerial>

Haines-Young R, Fish R. 2009. Hypothesis Testing. In Kitchin R, Thrift N (eds) *International Encyclopedia of Human Geography*, Volume 5, pp. 264–270. Oxford: Elsevier.

ISBN: 978-0-08-044911-1

© Copyright 2009 Elsevier Ltd.

## Hypothesis Testing

**R. Haines-Young**, University of Nottingham, Nottingham, UK

**R. Fish**, University of Exeter, Exeter, UK

© 2009 Elsevier Ltd. All rights reserved.

### Glossary

**Asymmetry between Proof and Refutation** The proposition that while no number of observations that are consistent with a theory or hypothesis prove that the theory is true, one counter observation can show the theory is false.

**Epistemology** The methods by which we are able to 'gain' knowledge about something; epistemology is the study of the principles and practices of enquiry.

**Hypothesis** A logical consequence of a theory that can be tested in some way.

**Ideographic** The term used to describe the research tradition in geography that emphasized the study of areal differentiation and the uniqueness of place, as in regional geography.

**Nomothetic** The term used to describe the research tradition in geography that emphasized the need to search for regularities and generalities that are expressed in theories that link geographical phenomena.

**Null Hypothesis** A device used in statistical hypothesis testing. The null hypothesis is set up as a competing hypothesis to the one the research seeks to explore. The null hypothesis usually asserts that there is no difference between observed and expected outcomes in some sampling process.

**Ontology** The basis upon which we recognize something as knowledge; ontology is the study of what kinds of things exist and therefore what kinds of things can be known about.

**Theory** A set of beliefs or assumptions that we hold about the structure of the world and the way it behaves.

**Type I and Type II Errors** The distinction made in statistics between the mistakes made by rejecting a hypothesis when it is true, compared to accepting a hypothesis when it is false.

### Introduction

To test a hypothesis is one of the most important things a researcher can do. Indeed so important is it that many would claim that it is the thing that distinguishes scientific research from other forms of academic inquiry. Certainly hypothesis testing is one of the ways that human geographers have sought to understand their discipline as a science. Perhaps not surprisingly then, it is a technique that is not without controversy, for its

use registers with wider debates about the nature and purposes of geographical knowledge. This article discusses the logical framework in which all types of hypothesis testing within human geography can be viewed and, through this, reviews some of the statistical tools available to those working with quantitative data. We then inspect some of the wider philosophical and practical complexities surrounding the process of hypothesis testing in geographical research, and what these issues share with more qualitative approaches to the production of geographical knowledge.

### Historical Context

Hypothesis testing emerged in the 1950s as part of debates about the scientific bases of geographical research. To be understood as a scientific enterprise, it was claimed that geography should create knowledge applicable to a range of conditions and circumstances. At that time, 'anglophone geography' had been wedded to the idea of regional analysis, a process of inquiry that emphasized areal differentiation and uniqueness. For some geographers, one of the consequences of this interest in specificity – in idiographic study – was that geography was weak when it came to the generation of models about how the world in general operated. Although the rich accounts of regional character and distinctiveness were interesting, it was argued that the explanatory potential of geography should rest on uncovering regularities and communalities in how social processes came to function across space. In contrast, hypothesis testing emerged as part of an approach to geographical research expressly designed to inform, and be informed by, theory; one that would seek to put its work on a more scientific footing through a concern with generalization – or nomothetic study.

The process by which this argument about generalization was accepted by many geographers and with it the development of hypothesis testing as a key feature of human geography, has been termed the quantitative revolution. This transformation in approach was the basis for understanding geographical endeavor as a spatial science. The development was one whose possibilities and limits have since been subject to great critical comment by human geographers. Hypothesis testing is more than simply a research 'technique' in this respect. It is part of the changing history of ideas concerning what geographical knowledge should look like, and the claims we should seek to make of it.

## Hypothesis Testing: The Logical Framework

Put simply, to test a hypothesis is to confront a theory with some kind of data. Theories and hypotheses are often regarded the same kind of thing – they represent our beliefs or guesses about the way the world works. However, it is useful to make a distinction between them. It is a difference that is clearly brought out through the act of testing.

To confront a theory with data means to take some logical consequence of an idea and try to find out whether it is supported by the observations that we make about the real world. A logical consequence is something that must be the case if the theory is true. For example, it is widely accepted that human activities are fundamentally affecting the atmospheric system, and that this will lead to long-term changes in climate. If this is indeed the case, then a number of things follow logically from this theory. Detectable changes to the concentration of atmospheric carbon should track growth in use of fossil fuels by people. Future human action to reduce our carbon footprint should result in a reversal of the warming trend. These logical consequences are best regarded as hypotheses. They are specific propositions that we can in principle check out by collecting some evidence. This checking process is what hypothesis testing is all about.

As the twentieth-century philosopher Karl Popper pointed out, the important thing to note about the situation we face when evidence conflicts with our hypothesis is that, providing we judge the evidence to be correct, we must reject the hypothesis and the theory from which it was derived. Of course, evidence can be ambiguous or difficult to interpret, and so we must not be too quick to refute the theory. Nevertheless, if we ultimately feel that the data are reliable, then the hypothesis must be considered wrong. As researchers we are not prepared to accept theories and data about the world that conflict with each other, and still assert that we understand or have explained something. Finding ourselves in this situation we must redesign our theories so that different logical consequences follow which are supported by the data. Alternatively, we must try to think up altogether different theories to explain the world around us which are better supported by the evidence.

Popper placed particular emphasis on the act of conjecturing theories or hypotheses and then attempting to refute them because of what he called the “asymmetry between proof and refutation.” This is the logical principle that while no number of observations consistent with a theory or hypothesis can prove it to be true, just one counter observation can show it to be false. He used this idea to describe what he thought was the essential characteristic that distinguished science from all other

activities – namely, the act of developing theories that are testable in the sense that they have consequences that can be checked out by observations, and the willingness to reject them if we find them wanting. Whether we regard ourselves as scientists or not, the idea is also relevant to other areas of research, in that we try to resolve the conflicts between our general ideas about the way we think the world works and the evidence we gain from our experience about it. Either our general ideas are mistaken, or the evidence is misleading. Hypothesis testing is one of the ways we try to eliminate such ambiguity.

## Hypothesis Testing and Quantification

The desire to measure or to quantify is not something that is unique to science, but it is one of the characteristics most often associated with it. The reason why scientists and indeed other researchers try to measure things is that it sometimes allows them to test hypotheses more easily by expressing ideas precisely. This is perhaps best illustrated by looking at some of the tools that are available for testing hypotheses using statistical methods.

Hypothesis testing belongs to that part of statistics called ‘inferential statistics’, which is concerned with the grounds on which one might accept or reject a proposition. Most statistical tests operate in the same way, in that they try, in the simplest situations, to pose two opposing or counter ideas against each other and look for evidence that might suggest that one should be rejected. The procedure is rather like ‘trial by jury’, in which we begin by assuming the innocence of the accused. The prosecution makes its case, and we are only prepared to reject our assumption of innocence if the evidence against them is ‘beyond reasonable doubt’. In hypothesis testing in statistics, the initial proposition is called the null hypothesis and the strategy is to see if there are any good grounds for rejecting it in favor of some alternative idea.

For example, suppose that we are interested in the knowledge farmers have about diffuse pollution problems and the contribution that their farms might make. We might be interested in this because as environmental managers or policy makers, we want to decide if there is a case for making better information available to farmers to help them minimize their environmental impacts. Dairy farming is often thought to be a major source of pollution and so as decision makers we may want to investigate if the views and beliefs of dairy farmers are different to those of other farmers.

**Box 1** shows the analysis of some data collected from 75 farms in South West England, in an area of high conservation importance that is dominated by dairy, beef, and mixed farms. The data were collected by interviewing farmers and coding up some of the questionnaire material

**Box 1 Analysis of knowledge of pollution sources among farmers using cross tabulation**

These data were collected through interviews with farmers from 75 farms in South West England. The area is of high conservation importance and is dominated by dairy, beef, and mixed farms.

A contingency test was used to examine the frequency (counts) with which different types of farmers recognize different types of pollution as potentially arising from their farm. These tables show the responses for dairy farms versus nondairy units for the pollution sources 'sediment' and 'nitrates'.

In both cases our null hypothesis is that there is no difference between dairy farmers and other farmers in the extent to which they recognize the different types of pollutants potentially arising from their land.

Data for sediment issue:

Sediment Issue		What sort of farm are you?		Total
		Nondairy	Dairy	
Farmers do not recognize problem	Observed Count	43	8	51
	Expected Count	38.8	12.2	51.0
Farmers recognize problem	Observed Count	14	10	24
	Expected Count	18.2	5.8	24.0
Total	Observed Count	57	18	75
	Expected Count	57.0	18.0	75.0

**Hypothesis test:**

The expected counts are those that would arise if dairy farmers had the same responses as other types of farmers in the sample. The Pearson Chi-Square statistic ( $\chi^2$ ) which compares observed and expected frequencies has a value of 6.04, with one degree of freedom, the probability of exceeding this value by chance is less than 0.014 (i.e., 1.4 times out of 100). Thus, we reject the null hypothesis that there is no difference between dairy farmers and other farmers in the extent to which they recognize that sediment runoff might potentially be arising from their land.

Data for nitrate issue:

Nitrate Issue		What sort of farm are you?		Total
		Nondairy	Dairy	
Farmers do not recognize problem	Observed Count	45	12	57
	Expected Count	43.3	13.7	57.0
Farmers recognize problem	Observed Count	12	6	18
	Expected Count	13.7	4.3	18.0
Total	Observed Count	57	18	75
	Expected Count	57.0	18.0	75.0

**Hypothesis test:**

In this case the null hypothesis is that there is no difference between dairy farmers and other farmers in the extent to which they recognize nitrates as a potential pollutant arising from their land. For these data, the differences between the observed and expected counts are much smaller. The Pearson Chi-Square statistic ( $\chi^2$ ) is equal to 1.131, and with one degree of freedom, we would exceed this value to be exceeded by chance with a probability of about 0.29 (i.e., 29 times out of 100). Thus, we cannot reject the null hypothesis. Dairy farmers appear to be no different to other farmers with respect to nitrates.

so that it could be analyzed quantitatively. The farmers were asked about the types of pollution they recognized might be coming from their farm; these included sediment, nitrates, phosphates, pesticides, herbicides, and

plastics. In areas where dairy and beef farming is widespread, nutrient problems may arise through poor management of slurry and manures, or as a result of dirty water running off from farm buildings and yards; it is

particularly a problem in dairy units where animals are gathered frequently. Disinfectants in dairy units may also be a source of phosphorus. Sediment runoff may be associated with dairy and beef farming but we might expect it to be less likely to be a problem in areas such as these where grassland is the main cover type.

In **Box 1** we have tabulated the frequency with which dairy farmers thought that the different types of pollution might be generated from their farm compared to all the other types of farmers interviewed. Our initial or null hypothesis is that dairy farmers are no different to other farmers. As decision makers we want to see if there is any evidence that might cause us to reject this idea in favor of one that suggests that the dairy farmers have a different level of awareness about the pollution that might be arising from their farms. In **Box 1**, the statistical test that is being used (Chi-squared, or  $\chi^2$ ) simply compares for each pollutant category, whether the difference between the number of times (frequency) with which the dairy farmers recognized that they possibly caused pollution was different to that recorded for the other farm types. The logic behind the test is that we are trying to see if a difference between the 18 dairy farms and the others was likely to be due to chance alone, that is, the kind of variation that would arise if we drew a subsample of 18 farms from the group of 75 at random. If the difference in frequencies is large enough, then we might suspect that the dairy farmers are behaving differently.

The analysis in **Box 1** suggests that dairy farmers are no different from other types of farmers in the extent to which they believe their operations generate the types of pollution they were asked about, such as nitrate, except in the case of sediment, where dairy farmers more frequently cited this possibility than other types of farmers. In the case of sediment, the difference between the two groups was much larger than we might expect by chance alone.

The statistical tools used for hypothesis testing normally give us an estimate of the chance or probability associated with any difference we observe between groups. In the case of the responses to the question about sediment, the size of the difference observed is only likely to have come about 1.4 times in 100 if we had repeatedly sampled sets of 18 farms at random from the group of 75. On the grounds that such a difference in frequency of response as that observed is unlikely, we might be prepared to reject our null hypothesis that in their response to the sediment question, dairy farmers are no different to other farmers, in favor of the alternative hypothesis that they are not responding in the same way.

If we are to fully understand the nature of hypothesis testing in statistics, we need to consider further the logic that underlies the tools being used. What would be better

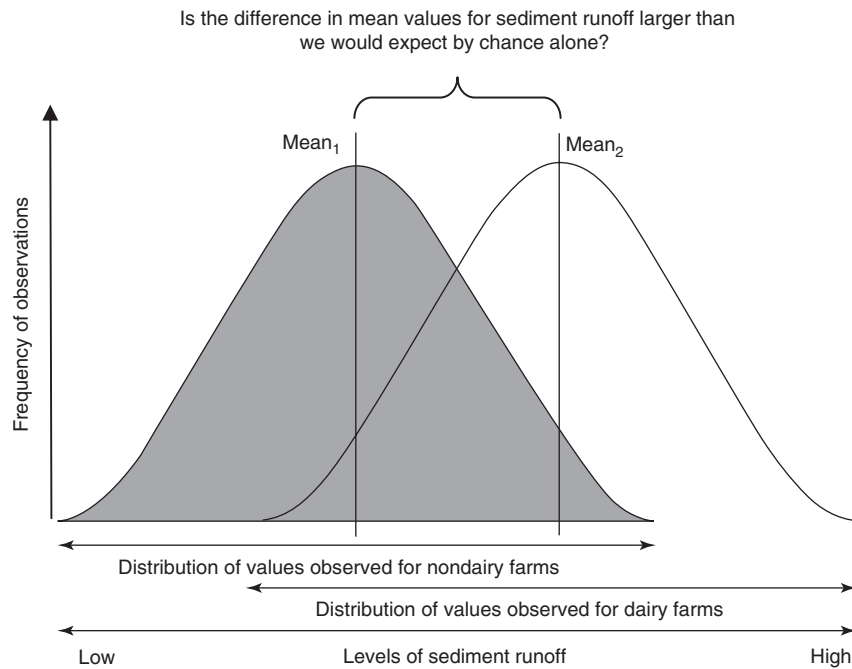
for a decision maker to do – to reject a null hypothesis when it is true, or to accept it when it is false? Statisticians refer to these two possibilities in terms of making ‘type I’ and ‘type II’ errors, respectively. In terms of our analogy between hypothesis testing and trial by jury, the first type of mistake is equivalent to finding an innocent person guilty, while the second represents the situation in which a guilty person goes free. Just as we think it better to risk letting guilty people go free in order to minimize the chances of an innocent person being wrongly convicted, so statisticians argue that we should exercise caution in our hypothesis testing. They suggest we should try to avoid making a type I error, by only rejecting a null hypothesis if there are ‘good grounds’ for doing so.

In our example concerning dairy farmers, we can minimize the chances of making a type I error if we are only prepared to reject the idea that they are no different to others if the chances of finding such a large difference between frequencies are small. Often people use the convention that we should be more than 95% certain, that is, the difference should arise no more than 5 times in 100. However, if it is really important that we avoid this type of error, then there is no reason why we cannot make the decision criteria much tougher – by being prepared to reject the null hypothesis if we are 99%, or even 99.9%, certain.

The example of statistical hypothesis testing presented here is a very simple one, and there are many more sophisticated tools available to test hypothesis using other types of data. For example, in our farm study, if we had measured sediment loss to rivers on the different types of farm, then a *t*-test could be used to investigate whether there was a difference in the mean levels observed between types (**Box 2**). Alternatively, if we had data on sediment loss over time following some awareness-raising program, we might ask whether there was any evidence of a decline in the rate of loss over time (**Box 3**). In each case, however, the approach is the same, the null hypothesis is set up to challenge the idea that we are putting forward, and we only reject it if there are good grounds for doing so. Moreover, when we do reject the null hypothesis we never claim that the alternative is true. Rather, we simply judge that there is reasonable evidence to investigate that alternative further.

### Contextualizing the Hypothesis-Testing Process

As the example shows, the process of hypothesis testing allows the researcher to start making some tentative claims about how the world operates. It also provides a framework in which the logic and assumptions underlying the reasoning process can be seen. Yet the act of hypothesis testing carries with it a wider set of

**Box 2 Testing the differences between potential levels of sediment runoff between farm types**

In this situation, a statistical technique called a *t*-test could be used to test the null hypothesis that there was no difference in the means of the daily sediment runoff values observed over, say one year, between dairy and nondairy farms. Only if the difference is larger than we would expect by chance alone would we reject the null hypothesis.

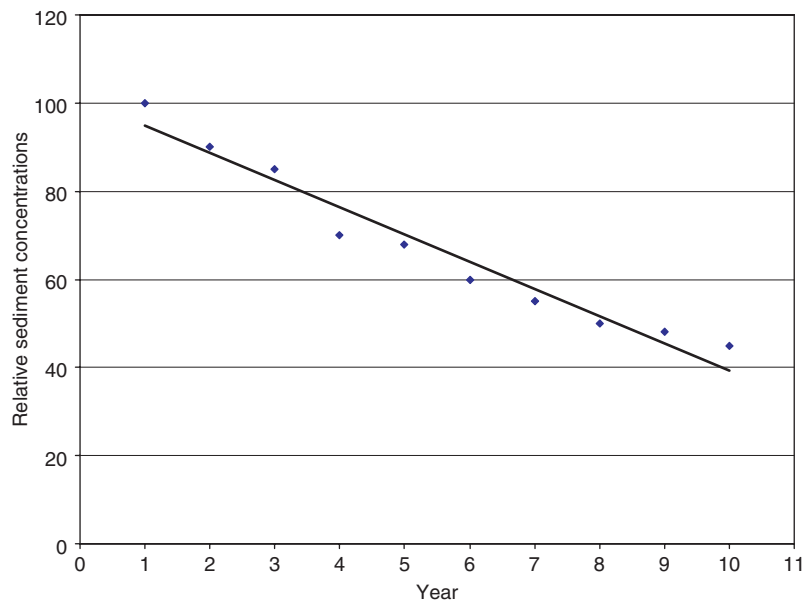
philosophical, practical, and ethical issues that affect how we evaluate it as a way of doing human geography research.

A good way of distinguishing between the issues at stake is to examine the process of hypothesis testing through the ideas of epistemology and ontology. By epistemology we mean the methods by which we are able to 'gain' knowledge about something. In this sense the use of hypothesis testing is an epistemological strategy. It is a technique by which the human geographer can begin to render the world meaningful in a certain kind of way, such as by exploring the association between farming sectors and attitudes toward diffuse pollution. In contrast, by ontology we mean the basis upon which we recognize something as 'knowledge' in the first place. In this sense hypothesis testing is not just a technique to produce knowledge about the world, but a means by which we explore 'what can be known'. This is true of any research process, whether based on hypothesis testing or not. Whichever epistemological strategy researchers have chosen to employ, they have already accepted, by implication, an ontological position. For those employing hypothesis testing, a very specific ontological claim is being made: namely that there are underlying patterns to the world that exist independently of our research into them, such as a potential relationship between farming

sector and understandings of diffuse pollution. Hypothesis testing therefore works on the promise that our efforts may begin to uncover these hidden regularities.

Confronting the act of hypothesis testing with the idea of epistemology raises interesting questions about how and why we go about researching the world in particular ways, and to what extent researching social processes can provide us with the type of information demanded by hypothesis testing techniques. Central to these is the epistemological assumption of objectivity. It is suggested that the creation of knowledge under hypothesis testing involves the researcher performing his or her task in a neutral or disinterested fashion. After all, if the world has regularities and patterns that exist independently of human observation, then it might be argued that the assumptions or values of the researcher are largely unimportant. Arguably anyone could arrive at identical conclusions if the same hypothesis testing procedure is followed.

The point we wish to emphasize here is that, as an epistemological strategy, hypothesis testing is part of an active human process involving judgments about the selection and framing of hypotheses, the assigning of significance to them, and the drawing of conclusions from results. Indeed, hypothesis testing must be understood as an instrument of research through which a range of

**Box 3 Testing hypotheses about changes in the levels of sediment runoff over time**

In this situation, a statistical technique called regression could be used to test the null hypothesis that there was no association between the yearly means of daily sediment runoff generated by dairy farms. Only if the slope of the line differed significantly from zero (that is the slope that we would expect if there is no association between the two variables) would we reject the null hypothesis of no change over time.

assumptions about the world can flow. This is apparent in a number of ways.

At one level, judgments arise when a researcher is faced with formulating a hypothesis. If it is one thing to set up a hypothesis that tests the relationship between farmer sectors and understandings of diffuse pollution, it is quite another to explain why the researcher would choose to pursue this hypothesis in the first instance. One answer to this is to say that these choices are logical consequences of a wider universe of hypotheses that contextualize the aims of a given piece of research (e.g., dairy farmers have different beliefs about pollution than other sectors). However, such a rationale does not explain why we choose to assign importance to diffuse pollution and agricultural sectors as a research issue.

Nor is the way we frame a hypothesis 'given' within a particular problem context. Even if the researcher could make a case for exploring the relationship between farming practices and pollution, this does not settle why the researcher would choose to examine attitudes as an important aspect of that debate. Researchers may choose to create and test hypotheses around a variety of equally plausible concerns, such as the market organization of different agricultural sectors, or the level of state assistance afforded to them. While it may be argued that the process of selecting and pursuing a given hypothesis is precisely designed to start discriminating between

contrary, yet plausible, understandings of given problem (i.e., to work with multiple hypotheses), it is nonetheless the case that by choosing one hypothesis over another a particular problem context tends to be emphasized as the way in which we think about research issues (e.g., 'attitudes rather than money are important to overcoming diffuse pollution'). Realistically, not all hypotheses will be explored and not all associations between relevant variables will be discerned. Like any other research processes, then, the researcher who employs hypotheses has to be very explicit about why and on what terms the questions are being framed. There is nothing objective about hypothesis testing in this respect. Even in the decisions about the level of statistical significance that we will apply in our hypothesis testing we make a moral judgment: how much certainty do we need before we start regulating the dairy farming sector?

Issues of epistemology and hypothesis testing also concern the terms on which data are collected and used to make claims about the world with some degree of confidence. The example we have used in this article is instructive for it uses a data-gathering technique widely employed by human geographers: that of the standardized, closed, and face-to-face interview. Thus, before we even start to look at associations between variables in the data set, we have to accept the notion that data we require will be both robust and collectable. Principally,

when conducting a survey like this we assume that interviewees understand what is meant by a particular question; answer a question when it is addressed to them; say what they mean when they respond to a question; say what they believe when offering responses; and respond in the terms that are expected. Researching humans as the basis for hypothesis testing is like any other research strategy in this respect: it comes with few guarantees. In the example data set we draw upon, we were fortunate to have 75 complete survey returns. But these returns in no way carry with them the certainty that we have comprehension, clarity, and frankness when making sense of the data. For example, why is it that dairy farmers recognized sediment as a problem but not the other types of pollution? Is it simply that it is more obvious or is sediment really an issue? Alternatively are farmers really telling us what they know about the other sources? Thinking about such issues is least one of the reasons why geographers researching human attitudes and values often emphasize the need to employ mixed methodologies, that link hypothesis testing to more probing and qualitative forms of investigation. What the example of dairy farmers shows, is that while superficially we appear to be testing only some simple, quantitative hypothesis, we are in fact confronting a whole network of hypotheses and assumptions that all need to be looked at critically before any final judgment can be made.

## Conclusion

Although many human geographers do not consider discussions of scientific methods to be relevant, or the use of statistical methods as being appropriate (to the kind of geography they do), whenever we seek to examine a case for something by reference to evidence, we are involved in the act of hypothesis testing – whether we formally acknowledge it or not. It can be argued that hypothesis testing in its broadest sense is therefore relevant to a very large part of what human geography does and we ought to be aware of what it involves. At the same time, at the epistemological level, those who test hypotheses as part of the research process can learn usefully from other, more qualitative and interpretative, forms of human geography inquiry that have encouraged researchers to consider more explicitly their positionality when choosing, applying, and drawing conclusions from hypotheses, and in developing research strategies that recognize the limits of the data collection process. One way to address these limits, and therefore build confidence in the research process, is to use hypothesis testing and quantitative inquiry as part of a larger, more pluralistic, research methodology.

Even so, it must be recognized that dealing with these epistemological dimensions of hypothesis testing the

ontological assumptions that guide its use *per se* are not being questioned. From the perspective of ontology, hypothesis testing has often been set apart from other arenas of human geography. Indeed, in light of its entrance into human geography under the quantitative revolution, many have since argued that the search for underlying patterns is problematical in the context of social systems, for human attitudes and behavior are like a moving target. That is to say, there is no amount of available data that can make these kinds of systems stable to the degree that hypothesis testing seems to aspire to. One response to this has been to suggest that the ontological claims underlying hypothesis testing need, at the very least, to be remodified to suit the human conditions they seek to account and speak for. Instead of regarding the world as a set of social processes whose rules can be progressively tamed and revealed, insights should instead aim to be ‘soft’ assessments, based on a more iterative understanding of what appears to ‘work’ in any given moment or circumstance. The extent to which such a situation precludes any requirement of researchers to test their social understandings in a critical way is a debate that must be had elsewhere.

See *also*: Critical Rationalism (After Popper); Philosophy and Human Geography; Physical Geography and Human Geography; Redbrick University Geography in Britain; Scientific Method.

## Further Reading

- Chalmers, A. F. (2004). *What's This Thing Called Science?* (3rd edn.). Buckingham: Open University Press.
- Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.
- Johnston, R. and Sidaway, J. (2004). *Geography and Geographers—Anglo-American Geography Since 1945*. London: Hodder.
- McGee, B. (1973). *Karl Popper (Modern Masters)*. New York: Penguin (Non-Classics).
- Peet, R. (1998). *Modern Geographical Thought*. Oxford: Blackwell.
- Policy Hub (2008). How research and evaluation evidence contributes to policy making. [http://www.policyhub.gov.uk/evaluating\\_policy/how\\_res\\_eval\\_evid.asp#philosophical](http://www.policyhub.gov.uk/evaluating_policy/how_res_eval_evid.asp#philosophical). (updated 12/2/2008)
- Popper, K. R. (1972). *The Logic of Scientific Enquiry (Sixth Impression)*. London: Hutchinson.
- Rogerson, P. A. (2001). *Statistical Methods for Geography*. London: Sage.
- Sayer, A. (1992). *Method in Social Science: A Realist Approach*. London: Routledge.

## Relevant Websites

- <http://ccnmtl.columbia.edu>  
Quantitative methods in the social sciences, QMSS e-lessons, Columbia Center for New Media Teaching and Learning.
- <http://www.eeng.dcu.ie/~tkpw>  
The Karl Popper Web.