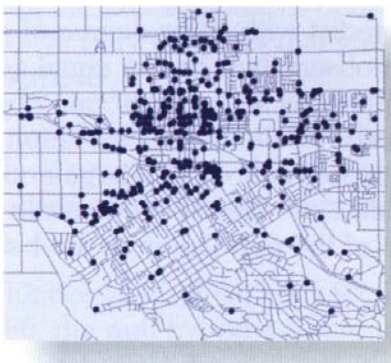
Testing statistical significance

The statistics in chapters 3, 4, and 5 allow you to identify a probable pattern or relationship. Before making a decision that involves committing resources or before doing further analysis, you'll want to know with some degree of certainty that your conclusions about the pattern or relationship are correct. Significance tests give you a probability that what a statistic is telling you is true. Probability is a measure of chance, and underlying all statistical tests are calculations that assess the role of chance on the outcome of your analysis.



The assaults look clustered, but you don't know if the clustering is due to chance, unless you test the statistical significance of the pattern.

STATISTICAL SIGNIFICANCE, PROBABILITY, AND SAMPLING

Statisticians distinguish between analyzing all possible observations within a study area (the population) and analyzing a sample. In the social sciences, since you can't talk to every person in a large area, predictions or assumptions are made about a group of people by collecting information from a few. In the natural sciences, samples are collected when you can't observe every individual. For example, when studying the distribution of plants in a region, it would be difficult to identify and note the location of every single plant of a particular species, so you'd set up a number of transects across the region and sample the plants along each transect. You'd then make estimates about the distribution of all the plants in the region from your sample.

The goal is to have the results of analysis for the sample match as closely as possible the results you would get had you been able to analyze the entire population. When a sample is a good representation of the population as a whole, you can expect measures for the sample to be near those of the population. Any difference you find is called sampling error. How well the sample matches the population depends on the characteristics of the sample, its size, and how it was collected.

The sample you collect is only one of many possible samples. If you obtained a different sample, you might very well draw a different conclusion about the population. Most samples—if they're large enough—would produce a good estimate of the characteristics of the population. Now and then you would, by chance, draw a sample that produced results very different from the true results for your study area; but the probabilities of drawing a highly unusual sample set would be very slim.

The smaller the sample, the more likely that one or a few extreme values included by chance will skew the results. With a larger sample, any extreme values that happen to be included in that particular sample will have a smaller effect on the distribution.

To predict with some degree of certainty, inferential statistics require that each observation has an equal and independent chance for inclusion in a sample—the selection of one observation should in no way influence the selection of the next observation. Samples that meet this requirement are said to be random. If a large enough random sample is collected, you can calculate the probability that the sample you've obtained is a good estimate of the population as a whole.

Confidence levels and null hypotheses

Statistical theory provides methods to measure levels of confidence associated with observations you make. The observations or beliefs are stated as a hypothesis. It's only human to favor any patterns or relationships you see or expect to see. So, to maintain impartiality, you set out to prove the opposite—the so-called null hypothesis. Your initial hypothesis is called the alternative. Significance tests help you decide whether you should or should not reject the null hypothesis.

In order to decide whether to reject the null hypothesis, you first decide the risk you are willing to accept for being wrong. This degree of risk, often referred to as the confidence level (or significance level), is expressed as a probability ranging from 0.0 to 1.0. The desired level of confidence is compared with an observed level of confidence to decide whether to reject or not to reject the null hypothesis. The observed level of confidence, known as the p-value, is calculated using your sample data. So the characteristics of your sample (its size, whether it's a random sample) affect the observed level of confidence.

The most common confidence levels for statistical tests are 0.10, 0.05, and 0.01. These values indicate the risk you are willing to accept for erroneously rejecting the null hypothesis. If the study could be repeated 100 times, each time with a different sample, probability indicates that, at the 0.05 confidence level, 95 out of 100 studies would yield accurate results. In other words, 5 out of 100 would likely yield erroneous results due to sampling error.

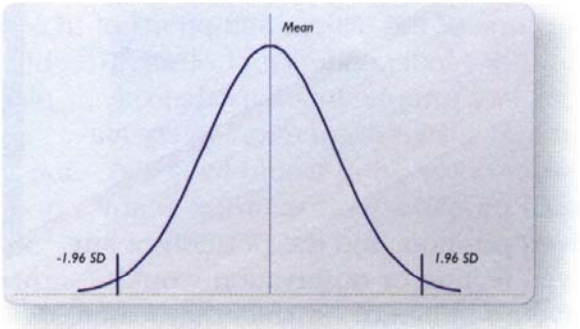
Each statistical tool has an appropriate significance test (sometimes more than one). The test provides a statistic that represents the p-value. And, the desired confidence level has a corresponding critical value (which depends on the specific test). If the value of the test statistic exceeds the critical value, you reject the null hypothesis. The results of your analysis are said to be statistically significant at the specified confidence level.

Rather than saying they accept the null hypothesis, statisticians say that they "don't reject" it. That's because you're never absolutely sure (it's all about probabilities), so you're not in a position to absolutely accept anything. Further analysis or new information may show that you should reject the null hypothesis. Similarly, rejecting the null hypothesis doesn't prove that the alternative hypothesis is true—it only indicates that your sample data does not support the null hypothesis.

Most spatial statistics tools calculate a test statistic at the same time they calculate the initial statistic and report both. Many of the tools calculate a Z-score, a reference measure for the standard normal distribution (with mean of zero and standard deviation of 1).

The critical value for the Z-score at a confidence level of 0.05 is 1.96. If the Z-score is within the range −1.96 to 1.96, the null hypothesis cannot be rejected. If it falls outside this range, you can reject the null hypothesis.

The critical values of −1.96 and 1.96 are standard deviations from the mean. Ninety-five percent of the area underneath the standard normal curve falls between plus and minus 1.96 standard deviations from the mean. The other 5% of the area is termed the rejection region. If the Z-score falls within the rejection region, there's only a 5% chance that you'd be wrong to reject the null hypothesis.



The critical values for Z at a confidence level of 0.05 are plus and minus 1.96 standard deviations from the mean—95% of the area under the curve (representing 95% of all value) falls within these limits.

At a confidence level of 0.01, 99% of the area under the curve falls between −2.58 and 2.58 standard deviations of the mean.

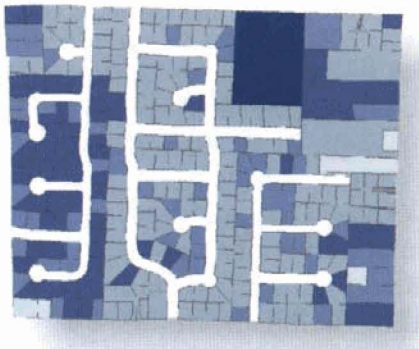
Confidence level	Z-score critical values
0.01	±2.58
0.05	±1.96
0.10	±1.65
0.20	±1.28

USING SIGNIFICANCE TESTS WITH SPATIAL DATA

These tests were developed within classic (nonspatial) statistics. Spatial data, however, contradicts some of the assumptions of inferential statistics, by its very nature. You need to be aware of the limitations of the tests, and the assumptions the tests make, and question the results. The result of the significance test—as is the statistic you're testing and, indeed, as is a map—is just one piece of information to aid in making a decision.

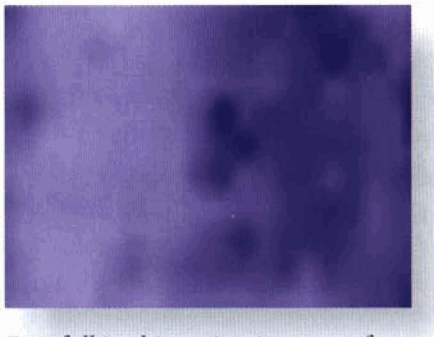
The tests are based on estimating the probability that the variation in your sample reflects the variation in the population as a whole. And they assume you're testing a random sample. In GIS analysis, though, it's often the case that the data has already been collected and put into a database. You may not know if the data was randomly sampled or how large the sample is compared to the population; you may not even know if you are dealing with a sample or with the whole population.

Even if you assume you are analyzing a sample, spatial data often violates one of the main assumptions of statistics—independence of observations in a sample. In a hypothetical random spatial distribution, every feature or observation would have an equal probability of occurring at any given position, and the position of any given feature or observation would have no influence on any other feature or observation in the dataset. But, the selling price for one particular house, for example, does in fact influence nearby housing values.



High-value properties (dark blue) are likely to be found together.

Similarly, the observations are often not random—commercial burglaries occur only where there are businesses, and businesses tend to cluster, so the outcome is somewhat predetermined. Also, spatial data is rarely evenly distributed across a region—for example, rainfall may increase as you move from west to east.



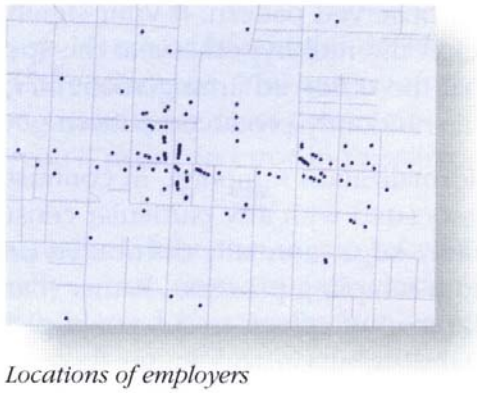
Rainfall in this region increases from west to east.

Finally, it's not uncommon in the GIS setting to find yourself working with very large datasets. Obtaining statistical significance in these cases will not be difficult if you're analyzing several thousand features or more.

Increases in computing power are allowing researchers to explore methods that rely on computer-generated distributions derived from millions of permutations over a dataset, such as Monte Carlo simulation and Bootstrapping. These methods, which may eventually replace theoretical distribution models and their accompanying mathematical equations, avoid the issues associated with statistical tests based on sampling (such as assumptions of normality or spatial independence). Nonetheless, the Z-score test and other tests based on sampling are still commonly used to assess statistical significance for spatial data.

SPATIAL DATA AND THE NULL HYPOTHESES

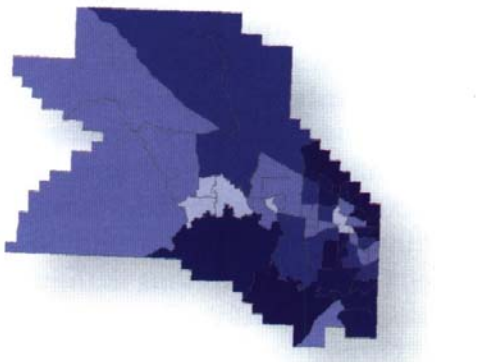
The null hypothesis for spatial pattern analysis is that the features are evenly distributed across the study area. However, in many cases, it's difficult to envision a situation in which the null hypothesis could or would ever be true. Suppose, for example, you have urban employment data and want to measure whether or not the locations of employment opportunities are clustered—it would be difficult to imagine an urban environment in which employment opportunities didn't cluster.



Locations of employers

While performing a statistical test to show that your employment data exhibits statistically significant clustering would not tell you anything you didn't already know, it would confirm your observations and understanding of what's happening. That could be important if decisions made using the analysis have legal or economic ramifications. Probably even more useful, though, would be to see if any strong clustering that is also statistically significant increases or decreases over time. That would tell you if employment is becoming more or less concentrated, and would provide insights into structural changes taking place in the city.

If you're analyzing the distribution of feature values, such as the median house value by census tract in a county, you have to make some assumptions about how the values were sampled (since the significance tests assume you're working with sample data). Two common sampling assumptions are randomization and normalization.

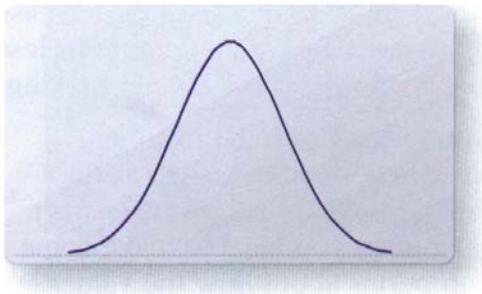


Census tracts for a county, color coded by median house value

Randomization sampling assumes that the observed spatial pattern of your data represents one of many possible spatial arrangements—the number of features having a particular value is always going to be the same (based on the observed number of each), but the arrangement can change. Suppose you have the number of cases per census tract of a disease. Some tracts have several cases of the disease; many don't have any. Your null hypothesis would be that the disease strikes randomly. If you could take the case count values for your study area and scatter them on a map of the census tracts, making sure every census tract got a value, you'd create a random pattern. The randomization null hypothesis postulates that if you could perform this operation infinite times, most of the time you would produce a pattern that was not markedly different from the observed pattern. If your significance test indicates that you should reject the null hypothesis at the specified confidence level, then you know that the observed arrangement of values would significantly differ from this randomly produced pattern.

Normalization sampling, in contrast, assumes that the number of cases associated with any particular census tract could be derived from an infinitely large, normally distributed population of values (through some random sampling process). Rather than scattering the observed values on the map of census tracts, you'd pick values from this hypothetical normal distribution and scatter those values on the map to create the random pattern.

With the randomization null hypothesis, you know up front all possible values; with the normalization null hypothesis, you don't know all possible values—you assume that the values are a sample from a larger population. That's hard to conceptualize, especially when you're analyzing spatially continuous data or data summarized by contiguous areas. If your study area is a county, for example, it would be hard to make the case that the set of values for all census tracts in the county is a sample.



Frequency curve for a normal distribution

The normalization null hypothesis not only assumes that your data is a sample, but also that the sample was obtained randomly and that the population from which the sample was obtained has a normal distribution of values. Every time you make an assumption about the data or the sample, you're potentially introducing error into the test. Randomization makes fewer assumptions than normalization, so it's safer to use, unless you know for sure your data matches the assumptions of normalization.

The Z-score (or other test result) will be calculated differently depending on the assumption, and hence the resulting value will be different. Since you compare the Z-score value to a critical value at a given significance level, the assumption you use has a bearing on whether the result is significant or not at that level.

Some software calculates and presents the results using both assumptions, in which case you'll need to decide which assumption best matches your situation, and use that result. In other cases, you may need to specify the assumption to use for the test statistic.

Reducing risk

You don't truly know if the null hypothesis is true or false—you decide to either reject it or not, with some level of confidence. You can try to reduce the risk that you'll draw the wrong conclusion from the significance test by minimizing the likelihood you'll make an error in rejecting the null hypothesis.

If you reject the null hypothesis and it actually is false, or—again based on your test—you don't reject it and it actually is true, then your conclusion is correct. But if you reject the null hypothesis and it's actually true, or you don't reject it and it's actually false, you're making an error. Statisticians call the first one a Type I error, and the other a Type II.

The risk is usually less with a Type I error. Suppose, for example, you're trying to identify potential environmental factors that may be contributing to a particular type of cancer. Your null hypothesis states that these cancer incidents are distributed randomly throughout your study area. If you commit a Type I error (you reject the null hypothesis, falsely concluding that cancer incidents are clustered), you will likely move to a more specific level of analysis, such as examining relationships among the cancer clusters and particular environmental factors, or seeing if you can reproduce your findings in a second study area. Hopefully, upon further scrutiny, your error will be uncovered. On the other hand, if you commit a Type II error (you fail to reject the null hypothesis, falsely concluding that this particular type of cancer strikes randomly), you may abandon your research prematurely. The bigger risk in this case lies with committing a Type II error.

Because with spatial statistics the null hypothesis is already established, the best you can do to favor a Type I error is specify a less stringent confidence level. If you use a confidence level of 0.1 instead of 0.05, you're more likely to say there's significant clustering (that is, reject the null hypothesis). If there actually isn't clustering (the null hypothesis is true), at worst you've only committed a Type I error. If you set a higher confidence level (say 0.01), you're less likely to say there's significant clustering (that is, you're more likely to accept the null hypothesis). If there is in fact clustering, you've committed a Type II error.

While statistical significance gives you some confidence that you have found a pattern or a probable relationship, it's not the final answer. There may be larger questions that factor into any decisions you make using the results of your analysis, such as whether your findings have important implications or even whether you're asking the right question in the first place. Regardless of the results of your test, you'll want to consider these larger issues when assessing the results of your analysis.

References

Bailey, Trevor C., and Anthony C. Gatrell. *Interactive Spatial Data Analysis.* Longman, 1995.

Burt, James E., and Gerald M. Barber. *Elementary Statistics for Geographers.* Guilford Press, 1996.

Earickson, Robert J., and John M. Harlin. *Geographic Measurement and Quantitative Analysis.* Macmillan, 1994.

Ebdon, David. *Statistics in Geography.* Blackwell, 1985.

Goodchild, Michael F. *Spatial Autocorrelation.* Catmog 47, Geo Books, 1986.

Griffith, Daniel A., and Carl G. Amrhein. *Statistical Analysis for Geographers.* Prentice-Hall, 1991.

Hamilton, Lawrence C. *Regression with Graphics: A Second Course in Applied Statistics.* Brooks/Cole, 1992.

Lee, Jay, and David W. S. Wong. *Statistical Analysis with ArcView GIS.* Wiley, 2001.