⊕ esri   🔍   ☰

What is a z-score? What is a p-value?

ArcGIS Pro 3.0

In this topic

1. Confidence Levels
2. FDR Correction
3. The Null Hypothesis and Spatial Statistics
4. Additional Resources

Most statistical tests begin by identifying a null hypothesis. The null hypothesis for the pattern analysis tools (Analyzing Patterns toolset and Mapping Clusters toolset) is Complete Spatial Randomness (CSR), either of the features themselves or of the values associated with those features. The z-scores and p-values returned by the pattern analysis tools tell you whether you can reject that null hypothesis or not. Often, you will run one of the pattern analysis tools, hoping that the z-score and p-value will indicate that you can reject the null hypothesis, because it would indicate that rather than a random pattern, your features (or the values associated with your features) exhibit statistically significant clustering or dispersion. Whenever you see spatial structure such as clustering in the landscape (or in your spatial data), you are seeing evidence of some underlying spatial processes at work, and as a geographer or GIS analyst, this is often what you are most interested in.

The p-value is a probability. For the pattern analysis tools, it is the probability that the observed spatial pattern was created by some random process. When the p-value is very small, it means it is very unlikely (small probability) that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis. You might ask: How small is small enough? Good question. See the table and discussion below.

Z-scores are standard deviations. If, for example, a tool returns a z-score of +2.5, you would say that the result is 2.5 standard deviations. Both z-scores and p-values are associated with the standard normal distribution as shown below.
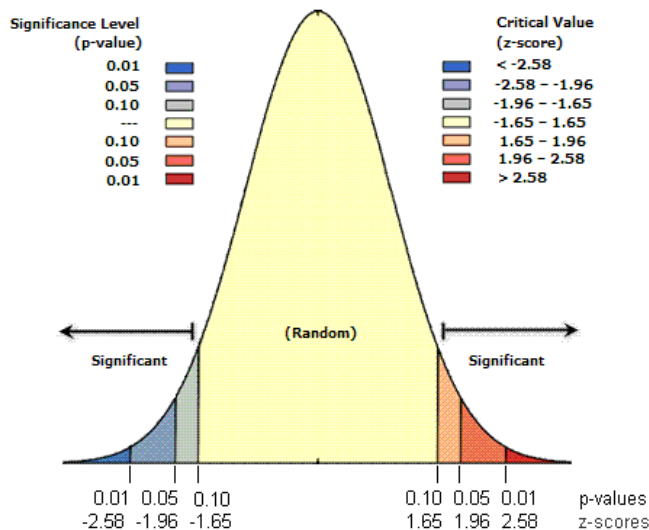
Very high or very low (negative) z-scores, associated with very small p-values, are found in the tails of the normal distribution. When you run a feature pattern analysis tool and it yields small p-values and either a very high or a very low z-score, this indicates it is unlikely that the observed spatial pattern reflects the theoretical random pattern represented by your null hypothesis (CSR).

To reject the null hypothesis, you must make a subjective judgment regarding the degree of risk you are willing to accept for being wrong (for falsely rejecting the null hypothesis). Consequently, before you run the spatial statistic, you select a confidence level. Typical confidence levels are 90, 95, or 99 percent. A confidence level of 99 percent would be the most conservative in this case, indicating that you are unwilling to reject the null hypothesis unless the probability that the pattern was created by random chance is really small (less than a 1 percent probability).

Confidence Levels

The table below shows the uncorrected critical p-values and z-scores for different confidence levels.

> **Note:**
> Tools that allow you to apply the False Discovery Rate (FDR) will use corrected critical p-values. Those critical values will be the same or smaller than those shown in the table below.

z-score (Standard Deviations)	p-value (Probability)	Confidence level
< –1.65 or > +1.65	< 0.10	90%
< –1.96 or > +1.96	< 0.05	95%
< –2.58 or > +2.58	< 0.01	99%

Consider an example. The critical z–score values when using a 95 percent confidence level are –1.96 and +1.96 standard deviations. The uncorrected p–value associated with a 95 percent confidence level is 0.05. If your z–score is between –1.96 and +1.96, your uncorrected p–value will be larger than 0.05, and you cannot reject your null hypothesis because the pattern exhibited could very likely be the result of random spatial processes. If the z–score falls outside that range (for example, –2.5 or +5.4 standard deviations), the observed spatial pattern is probably too unusual to be the result of random chance, and the p–value will be small to reflect this. In this case, it is possible to reject the null hypothesis and proceed with figuring out what might be causing the statistically significant spatial structure in your data.

A key idea here is that the values in the middle of the normal distribution (z–scores like 0.19 or –1.2, for example), represent the expected outcome. When the absolute value of the z–score is large and the probabilities are small (in the tails of the normal distribution), however, you are seeing something unusual and generally very interesting. For the Hot Spot Analysis tool, for example, unusual means either a statistically significant hot spot or a statistically significant cold spot.

FDR Correction

The local spatial pattern analysis tools including Hot Spot Analysis and Cluster and Outlier Analysis Anselin Local Moran's I provide an optional Boolean parameter Apply False Discovery Rate (FDR) Correction. When this parameter is checked, the False Discovery Rate (FDR) procedure will potentially reduce the critical p–value thresholds shown in the table above in order to account for multiple testing and spatial dependency. The reduction, if any, is a function of the number of input features and the neighborhood structure employed.

Local spatial pattern analysis tools work by considering each feature within the context of neighboring features and determining if the local pattern (a target feature and its neighbors) is statistically different from the global pattern (all features in the dataset). The z–score and p–value results associated with each feature determines if the difference is statistically significant or not. This analytical approach creates issues with both multiple testing and dependency.

Multiple Testing—With a confidence level of 95 percent, probability theory tells us that there are 5 out of 100 chances that a spatial pattern could appear structured (clustered or dispersed, for example) and could be associated with a statistically significant p–value, when in fact the underlying spatial processes promoting the pattern are truly random. We would falsely reject the CSR null hypothesis in these cases because of the statistically significant p–values. Five chances out of 100 seems quite conservative until you consider that local spatial statistics perform a test for every feature in the dataset. If there are 10,000 features, for example, we might expect as many as 500 false results.

Spatial Dependency—Features near to each other tend to be similar; more often than not spatial data exhibits this type of dependency. Nonetheless, many statistical tests require features to be independent. For local pattern analysis tools this is because spatial dependency can artificially inflate statistical significance. Spatial dependency is exacerbated with local pattern analysis tools because each feature is evaluated within the context of its neighbors, and features that are near each other will likely share many of the same neighbors. This overlap accentuates spatial dependency.

There are at least three approaches for dealing with both the multiple test and spatial dependency issues. The first approach is to ignore the problem on the basis that the individual test performed for each feature in the dataset should be considered in isolation. With this approach, however, it is very likely that some statistically significant results will be incorrect (appear to be statistically significant when in fact the underlying spatial processes are random). The second approach is to apply a classical multiple testing procedure such as the Bonferroni or Sidak corrections. These methods are typically too conservative, however. While they will greatly reduce the number of false positives they will also miss finding statistically significant results when they do exist. A third approach is to apply the FDR correction which estimates the number of false positives for a given confidence level and adjusts the critical p–value accordingly. For this method statistically significant p–values are ranked from smallest (strongest) to largest (weakest), and based on the false positive estimate, the weakest are removed from this list. The remaining features with statistically significant p–values are identified by the Gi_Bin or COType fields in the output feature class. While not perfect, empirical tests show this method performs much better than assuming that each local test is performed in isolation, or applying the traditional, overly conservative, multiple test methods. The additional resources section provides more information about the FDR correction.

The Null Hypothesis and Spatial Statistics

Several statistics in the Spatial Statistics toolbox are inferential spatial pattern analysis techniques including Spatial Autocorrelation (Global Moran's I), Cluster and Outlier Analysis (Anselin Local Moran's I), and Hot Spot Analysis (Getis–Ord Gi*). Inferential statistics are grounded in probability theory. Probability is a measure of chance, and underlying all statistical tests (either directly or indirectly) are probability calculations that assess the role of chance on the outcome of your analysis. Typically, with traditional (nonspatial) statistics, you work with a random sample and try to determine the probability that your sample data is a good representation (is reflective) of the population at large. As an example, you might ask "What are the chances that the results from my exit poll (showing candidate A will beat candidate B by a slim margin) will reflect final election results?" But with many spatial statistics, including the spatial autocorrelation type statistics listed above, very often you are dealing with all available data for the study area (all crimes, all disease cases, attributes for every census block, and so on). When you compute a statistic for the entire population, you no longer have an estimate at all. You

have a fact. Consequently, it makes no sense to talk about likelihood or probabilities anymore. So how can the spatial pattern analysis tools, often applied to all data in the study area, legitimately report probabilities? The answer is that they can do this by postulating, via the null hypothesis, that the data is, in fact, part of some larger population. Consider this in more detail.

The Randomization Null Hypothesis—Where appropriate, the tools in the Spatial Statistics toolbox use the randomization null hypothesis as the basis for statistical significance testing. The randomization null hypothesis postulates that the observed spatial pattern of your data represents one of many (n!) possible spatial arrangements. If you could pick up your data values and throw them down onto the features in your study area, you would have one possible spatial arrangement of those values. (Note that picking up your data values and throwing them down arbitrarily is an example of a random spatial process). The randomization null hypothesis states that if you could do this exercise (pick them up, throw them down) infinite times, most of the time you would produce a pattern that would not be markedly different from the observed pattern (your real data). Once in a while you might accidentally throw all the highest values into the same corner of your study area, but the probability of doing that is small. The randomization null hypothesis states that your data is one of many, many, many possible versions of complete spatial randomness. The data values are fixed; only their spatial arrangement could vary.

The Normalization Null Hypothesis—A common alternative null hypothesis, not implemented for the Spatial Statistics toolbox, is the normalization null hypothesis. The normalization null hypothesis postulates that the observed values are derived from an infinitely large, normally distributed population of values through some random sampling process. With a different sample you would get different values, but you would still expect those values to be representative of the larger distribution. The normalization null hypothesis states that the values represent one of many possible samples of values. If you could fit your observed data to a normal curve and randomly select values from that distribution to toss onto your study area, most of the time you would produce a pattern and distribution of values that would not be markedly different from the observed pattern/distribution (your real data). The normalization null hypothesis states that your data and their arrangement are one of many, many, many possible random samples. Neither the data values nor their spatial arrangement are fixed. The normalization null hypothesis is only appropriate when the data values are normally distributed.

Additional Resources

- Ebdon, David. *Statistics in Geography.* Blackwell, 1985.
- Mitchell, Andy. *The ESRI Guide to GIS Analysis,* Volume 2. ESRI Press, 2005.
- Goodchild, M.F., *Spatial Autocorrelation.* Catmog 47, Geo Books, 1986
- Caldas de Castro, Marcia, and Burton H. Singer. "Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Test in Local Statistics of Spatial Association." *Geographical Analysis* **38**, pp 180–208, 2006.

Related topics

- [High/Low Clustering (Getis–Ord General G)](#)
- [Spatial Autocorrelation (Global Moran's I)](#)
- [Cluster and Outlier Analysis (Anselin Local Moran's I)](#)
- [Hot Spot Analysis (Getis–Ord Gi*)](#)
- [Ordinary Least Squares (OLS)](#)
- [Optimized Hot Spot Analysis](#)
- [Emerging Hot Spot Analysis](#)