Glossary of terms

Glossary of terms

1. **Binomial model**: A statistical model used to analyse data with two possible outcomes, typically denoted as "success" and "failure."
2. **Covariate**: A variable that is not the primary focus of a study but is considered alongside the main variables to understand potential relationships or influences.
3. **Correlation**: A statistical measure that indicates the degree to which two variables change together, often expressed as a coefficient between -1 and 1.
4. **Data noisiness**: The presence of random or irrelevant variations in data, which can make it challenging to extract meaningful information.
5. **Design-based method**: A statistical approach that considers the specific design or structure of a study when making inferences from data.
6. **DHS (Demographic and Health Survey)**: A large-scale survey programme that collects data on demographic and health indicators in various countries.
7. **Elimination**: The reduction or eradication of a disease or condition from a specific area or population.
8. **Friction surface**: A spatial model representing the resistance or cost of moving between different locations in a geographic area.
9. **Geoconnect ID**: An identifier used to link or connect geospatial data to specific locations or regions.
10. **Georeferenced data**: Data that are associated with specific geographic coordinates, allowing them to be mapped and analysed in a geographic context.
11. **Geospatial**: Refers to data, information, or activities that have a geographic component and can be represented on a map.
12. **GLMM (Generalised Linear Mixed Model)**: A statistical model that extends the generalised linear model by incorporating random effects to account for correlation or hierarchical data structures.
13. **GTMP (Global Trachoma Mapping Project)**: A global initiative that ran between 2012-2015 aimed at mapping the prevalence of trachoma, a neglected tropical disease that causes blindness.
14. **Ground-truthing**: The process of verifying or validating data, typically by collecting on-the-ground information to confirm their accuracy.
15. **ITI (International Trachoma Initiative)**: An organisation dedicated to the elimination of trachoma as a public health problem.
16. **Landcover**: The physical and biological characteristics of the Earth's surface, including vegetation, water bodies, and built environments.
17. **Linear relationship**: A statistical relationship between two variables that can be represented by a straight line on a scatterplot.
18. **Logistic regression**: A statistical model used to analyse the relationship between a binary dependent variable and one or more independent variables.

19. **MBG (Model-Based Geostatistics)**: A statistical approach that combines spatial modelling and geostatistics to make predictions about data in unsampled locations.

20. **Model calibration**: The process of adjusting model parameters to ensure that the model's predictions align with observed data.

21. **Model distribution**: The probability distribution used to describe the variability of data in a statistical model.

22. **Model validation**: The process of assessing how well a statistical or predictive model performs by comparing its output to real-world data.

23. **Multivariate**: Involving multiple variables or factors, often used in statistical analysis to understand complex relationships.

24. **NTD (Neglected Tropical Disease)**: A group of infectious diseases that primarily affect people in tropical and subtropical regions, often with limited access to healthcare.

25. **PBT (Probability of being Below Threshold)**: A metric used to estimate the probability that the disease elimination target has been achieved.

26. **Population density**: The number of people living per unit area, often measured in individuals per square kilometre or square mile.

27. **Population weighting**: A technique used in statistical analysis to account for variations in population size when making inferences.

28. **R software**: An open-source programming language and software environment commonly used for statistical analysis and data visualisation.

29. **Random effect**: In a statistical model, a variable that represents random or unexplained variation, often accounting for correlation within hierarchical data.

30. **Raster file**: A data format that divides geographic information into a grid of cells, where each cell has a value representing a particular attribute.

31. **Shapefile**: A common file format for storing geographic vector data, such as points, lines, and polygons.

32. **Variance of the spatial process**: A measure of how data points vary in space, indicating the degree of spatial autocorrelation.

33. **Simulation study**: A research method that uses computer-generated data to model and analyse real-world phenomena.

34. **Spatial confounding**: A potential issue in spatial analysis where unaccounted factors lead to misleading results.

35. **Statistically significant**: A result in a statistical analysis that is unlikely to have occurred by random chance, often expressed with a significance level (e.g., p-value).

36. **Strata**: Divisions or subsets used to group data for analysis, often based on specific criteria or characteristics.

37. **TF (Trachomatous inflammation—follicular)**: A clinical sign of trachoma, indicating inflammation with follicles on the inside of the upper eyelid.

38. **Threshold**: A specific value or condition used to categorise or make decisions in data analysis.

39. **TT (Trachomatous Trichiasis)**: A late stage of trachoma, characterised by the in-turning of the eyelashes that tough the eyeball, which can lead to blindness.

40. **Univariate**: Involving a single variable or factor, often used in statistical analysis to describe or analyse data

with only one dimension.