Sources of Error (and other topics)

CSC Geographic Information Systems

13 November 2014

- sources of error
  - projections
    - earth is round, map is flat
    - distortion of shapes, angles, distances, and/or areas
  - imperfect datums / approximations of earth's shape
    - earth is not a sphere
    - earth is not even an ellipsoid!
  - sampling and interpolation
    - cannot measure everywhere!
    - must estimate at values of points between measured points
  - measurement errors
    - finite resolution of instruments
  - human error
    - transcription
    - miscalibrated instruments
    - lost data
  - finite precision of arithmetic on computers
- Interpolation and extrapolation

  interpolation values between measured points (do this more confidently)

  extrapolation values beyond range of measured points (less confidence when we do this)
- method of least squares
  - a way of identifying relationships among variables
  - could be used to help explain and predict phenomena
  - let's say each of many measurement has two components—e.g., we could measure the moisture in each square meter of a farmer's field and measure the amount of wheat harvested from each square meter

  - can plot data on x-y graph
  - suspect a linear relationship between variables—e.g,, $yield = m * moisture + b$
  - want to find m (slope—steepness of line) and b (intercept—height of the point at which the line intersects the y axis) that gives line that is "closest" to all data points
  - closest means sum of squared errors is least
  - minimize $\sum_0^N (y_i - (m \cdot x_i + b))^2$
  - best fit
  - could derive formula using our knowledge of calculus or use built-in functions in spreadsheet or statistical software
- statistical reasoning is important in work with GIS
- Different definitions of distance
  - Euclidean
    - "as the crow flies"
    - computed like length of hypoteneuse of right triangle
    - $distance = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$
  - Manhattan
    - like directions you would give to someone in a big city
    - "east 3 blocks, then north 6 blocks"
    - $distance = |x_1 - x_0| + |y_1 - y_0|$
  - common features of all distance functions
    - $distance(a, a) = 0$
    - $distance(a, b) = distance(b, a)$
    - $distance(a, c) \leq distance(a, b) + distance(b, c)$
  - other functions can also be generalized (e.g., average—there are arithmetic, geometric, and harmonic means)