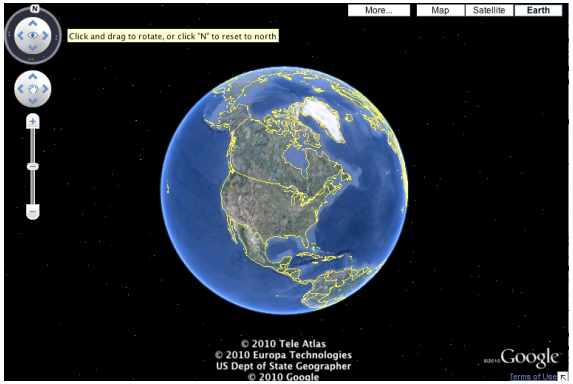

How Deep is the Ocean?

1 What percentage of the world's surface is covered by water?

The data provided by the Scripps Institution of Oceanography [see Oceanography Data on the link opposite the BIOS601 topic "Measurement, Surveys..."] can provide an answer, but some work is required on your part.

i. Form a simple random sample of 200 locations on the Earth's surface, and obtain from the SRTM30_PLUS database the land elevation / ocean depth at each of these these randomly selected locations. From these 'readings', calculate a point estimate of the percentage. Also calculate a (probabilistic) margin of error: do this by calculating a standard error, and multiplying it by say 1.96 so that you can make a probabilistic statement.

ii. Are you worried about the appropriateness of using 1.96 (and the Normal distribution) for 95% confidence? Why/why not?

iii. The root mean squared error includes both sampling variation and non-sampling errors. Your margin of error is limited to the sampling variation and does not include non-sampling errors. Give an example of one non-sampling error.

2 What is the average depth of the ocean?

i. From the subset of relevant observations (from among the entire 200), estimate the mean ocean depth, and calculate an accompanying margin of error. Even though there is a random component to it, pretend that the sample size was predetermined.

ii. Are you worried about the appropriateness of using 1.96 (and the Normal distribution) for 95% confidence? Why/why not?

3 Ensuring that a sample of n' locations will yield 200 [or more] usable ones

i. How big must n' be in order to have a good chance (say 80%) that it will yield at least 200 usable ones (i.e. ocean locations)?

ii. What if you sampled sequentially until, at the n'-th draw, you reached the 200-th usable one? What statistical distribution describes the random variable n' ? How could you calculate its 10-th and 90-th percentiles? (pretend you know the value of the parameter that determines its distribution).

4 More efficient (or more practical) sampling strategies

(Very briefly) describe the circumstances in which a sampling scheme other than s.r.s (systematic, stratified, cluster) would offer either practical or statistical efficiency advantages; mention also the downsides of these schemes.

5 Oh Oh

What if your research assistant spent all the research budget obtaining the data for a sample of 200 locations, but where the latitude locations were $\sim U(-90, 90)$ and likewise the longitude locations were $\sim U(-180, 180)$?

Are the data worthless?