Statistical inference; estimate, precision and bias

For application to GHG inventories, the IPCC defines two good practice criteria (Penman et al., 2003):

I. "neither *over- nor under-estimates* as far as can be judged"
II. "*uncertainties are reduced* as far as practicable"

Criterion I relates to concept of *bias* -- property of an *estimator* which, when applied to sample data, produces an *estimate*

Criterion II: estimate might deviate from true value -- a *confidence interval* expresses the uncertainty of a estimate

Real life example of inference

Votes for Obama vs. Romney, U.S. presidential election 2012?

Let's assume a simple random sample (SRS) of 500 voters; we implement a *SRS estimator* to get an *unbiased estimate* of votes:

$$\hat{\mu}_{Obama} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1+1+0+...+1}{500} = 0.51$$

Real life example of inference

A 95% *confidence interval* means that 95% of such intervals, one for each set of sample data, include the true value. The interval width is related to *precision*, a measure of the *uncertainty* addressed by IPCC criterion II. A CI is calculated as the product of the standard error and the z- or t-score.

$$\hat{V}(\hat{\mu}_{Obama}) = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu})^2}{n-1} = 0.003$$

$$1.96\sqrt{\hat{V}(\hat{\mu}_{Obama})} = 0.11$$

Inference in a geographic context

We can make inference of area of activity data -- in this case we use a sample of *reference observations* of the land surface instead of voters

A map is not required for estimation but a “good” map will *increase precision* of estimates

If the parameter of interest (e.g. deforestation) is a small proportion of the population (the map), a *stratification* will be very valuable

NOTE: we are not validating the map; and map accuracy is often of secondary importance! We want an unbiased estimate ± CI!

How do we know the SRS estimator is *unbiased*?

$$E(\overline{y}) = E(\frac{y_1 + y_2 + \ldots + y_n}{n}) =$$

$$= \frac{E(y_1) + E(y_2) + \ldots E(y_n)}{n} =$$

$$= \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$