Marine Resources Ressources marines

Statistical considerations for structuring your data collection programme

Artisanal Tuna Data Workshop
11^{th} – 14^{th} November 2013

Dr. Aaron Berger
OFP, SPC

CREEL SURVEY

The data collection programme can be described as a creel survey through the use of visual enumeration, fisher interviews, and catch inspection/measurements.

The fact that we can say we are setting up a fishery dependent, boat-based access type creel survey does not provide information on:

1. what data we need to collect (discussed earlier)
2. how to collect it (sample design)
3. how much is needed (sampling intensity).

Note, that information from the creel survey (e.g., # trips) can be used to ‘raise’ catch and effort information from logsheet submissions.

CREEL SURVEY

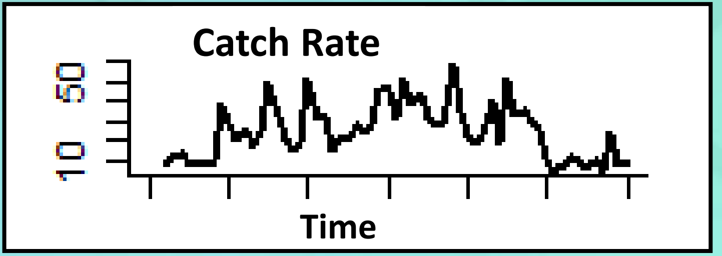
Creel surveys are often categorised by the method used to intercept the vessel. The type of creel survey to be used will depend on the local conditions and access to landing sites.

Access survey – The sampler stays at a fixed point
Roving Survey – The samplers moves between different sampling points
Bus route type access survey- The sampler moves to each location and sample for a fixed amount of time.
Aerial survey – The sampler flys over the 'sampling' area.

The data collection programme is an access creel survey - samplers are generally stationed at a fixed point – providing a "fishery dependent survey" as the data is gathered from a fishing activity (in contrast to research surveys).

WHY?

Estimate total catch by species and effort (obtain catch rates) to monitor population trends through time (long-term).



Logsheets provide the catch and effort for some subset of vessels.

What about the other 'non-sampled' vessels?

Need to 'infer' the effort and catch of those vessels. (statistical inference)

- Logsheets provide average CPUE (catch/day or catch/trip)
- Creel survey to estimate total effort (# of days or trips)

SAMPLING FRAME

A sampling frame records and preferably maps all of the fishery variables (i.e. number of boats, list of fishing gears, list of target species etc).

The sampling frame will show where the fisheries are replicated; that is landing sites or islands that have the same fishing variables (vessel types, fishing gear, target species etc).

The sample frame is used in the sampling design document. Once the fishery has been documented and mapped decisions can be made about where sampling can/should take place.

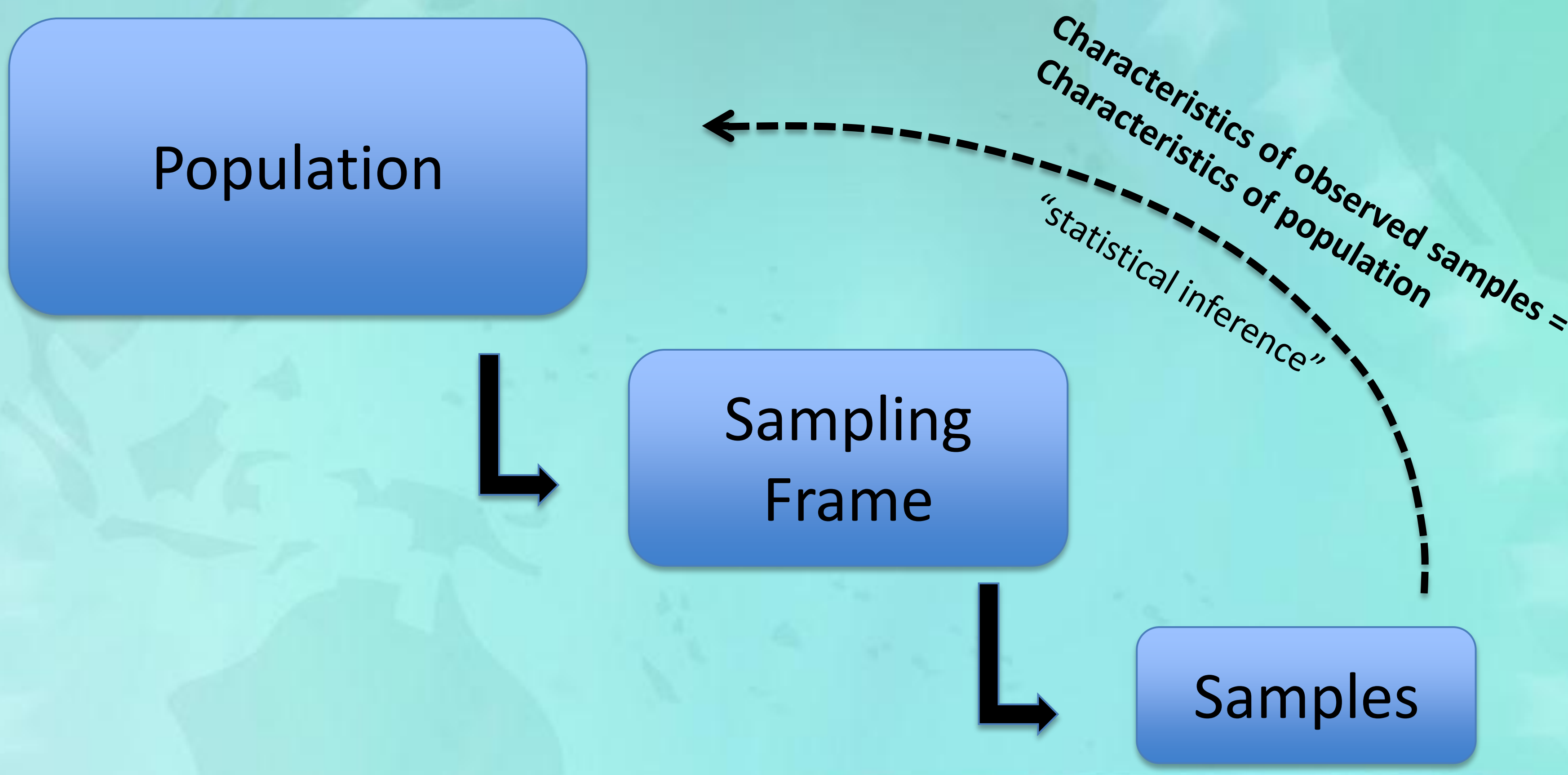
Generally, funding and the ability to access the fishery landing sites will often be the deciding factors when ultimately deciding the sampling design (not always 'the best' from a pure statistical standpoint - reality).

DEVELOPING A SAMPLING DESIGN

The sampling design should outline:

- The sampling frame – a list and a map of the fishing variables.
- The questions that the sampling programme will endeavour to answer. Noting that questions can change over-time. However, having good catch and effort data will go a long-way to answering any current or future questions.
- Highlight any unique fishing variables (i.e. vessel type: sail) so these can be included in the selection of sampling points, whenever possible.
- Show the selection of islands / landing points where access is possible with staff, transport and budget considerations.
- Show the random selection of the islands / landing points that will be sampling.
- Indicate whether the sampling programme will be long-term or short-term.

DEVELOPING A SAMPLING DESIGN



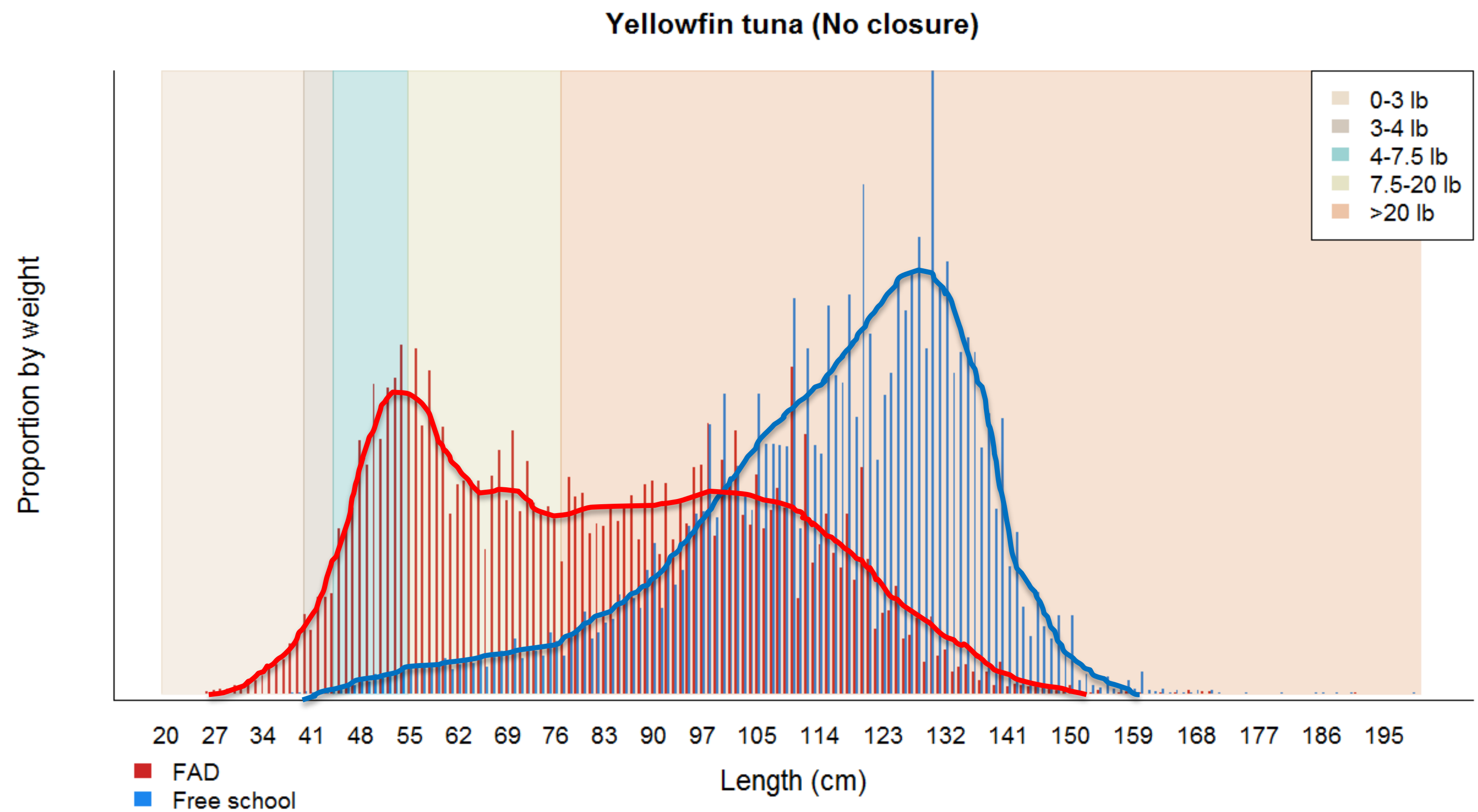
SAMPLING DESIGN

'Samples' should ideally be collected at random and be representative.

Random – by selecting samples at random the likelihood of biased estimates is drastically reduced. Probability of selecting each sample is the same.

Non-randomness introduces *bias* into sampling (inaccurate view of population)
Examples:

- sample only boats with lots of catch (reach sample quota quicker) – average catch per trip biased high
- visit only busy ports (# trips per day biased)
- get length frequency information from only one type of fishing gear


Yellowfin tuna (No closure)
0-3 lb
3-4 lb
4-7.5 lb
7.5-20 lb
>20 lb
Proportion by weight
20 27 34 41 48 55 62 69 76 83 90 97 105 114 123 132 141 150 159 168 177 186 195
Length (cm)
FAD
Free school


CATCH IT REPORT IT

SAMPLING DESIGN

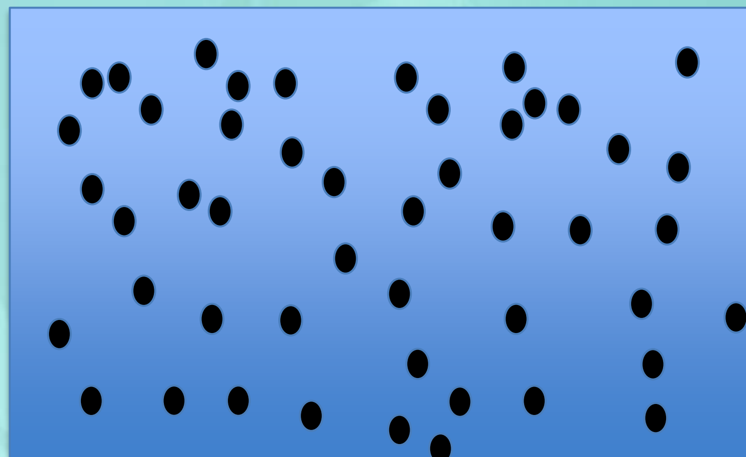
Representative – samples should reflect or represent the population under consideration.

For example – stratified random sampling

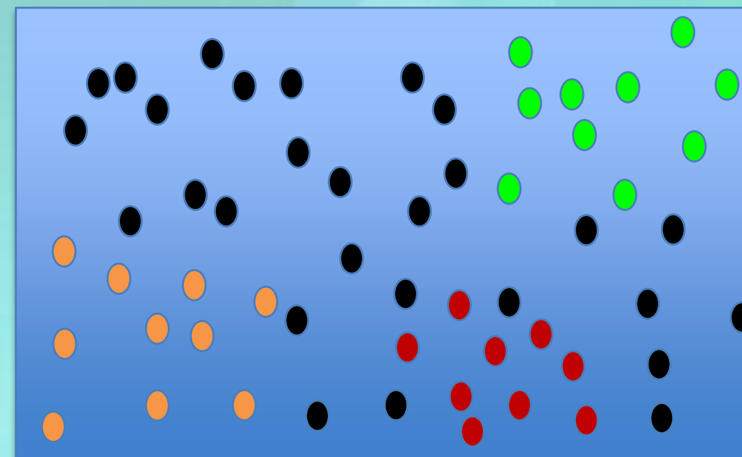
There may often be factors which divide up the population into sub-populations (strata) and we may expect the measurement of interest to vary among the different sub-populations. This has to be accounted for when we select a sample from the population in order to ensure that we obtain a sample that is representative of the population.

Important strata could be vessel size class, gear type, or port of unloading.

one strata

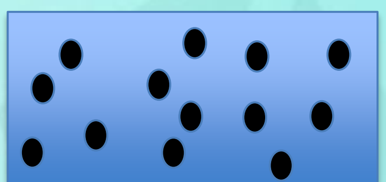


four strata

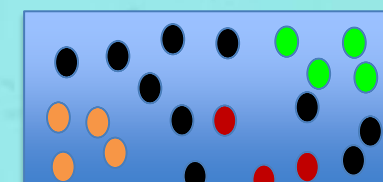


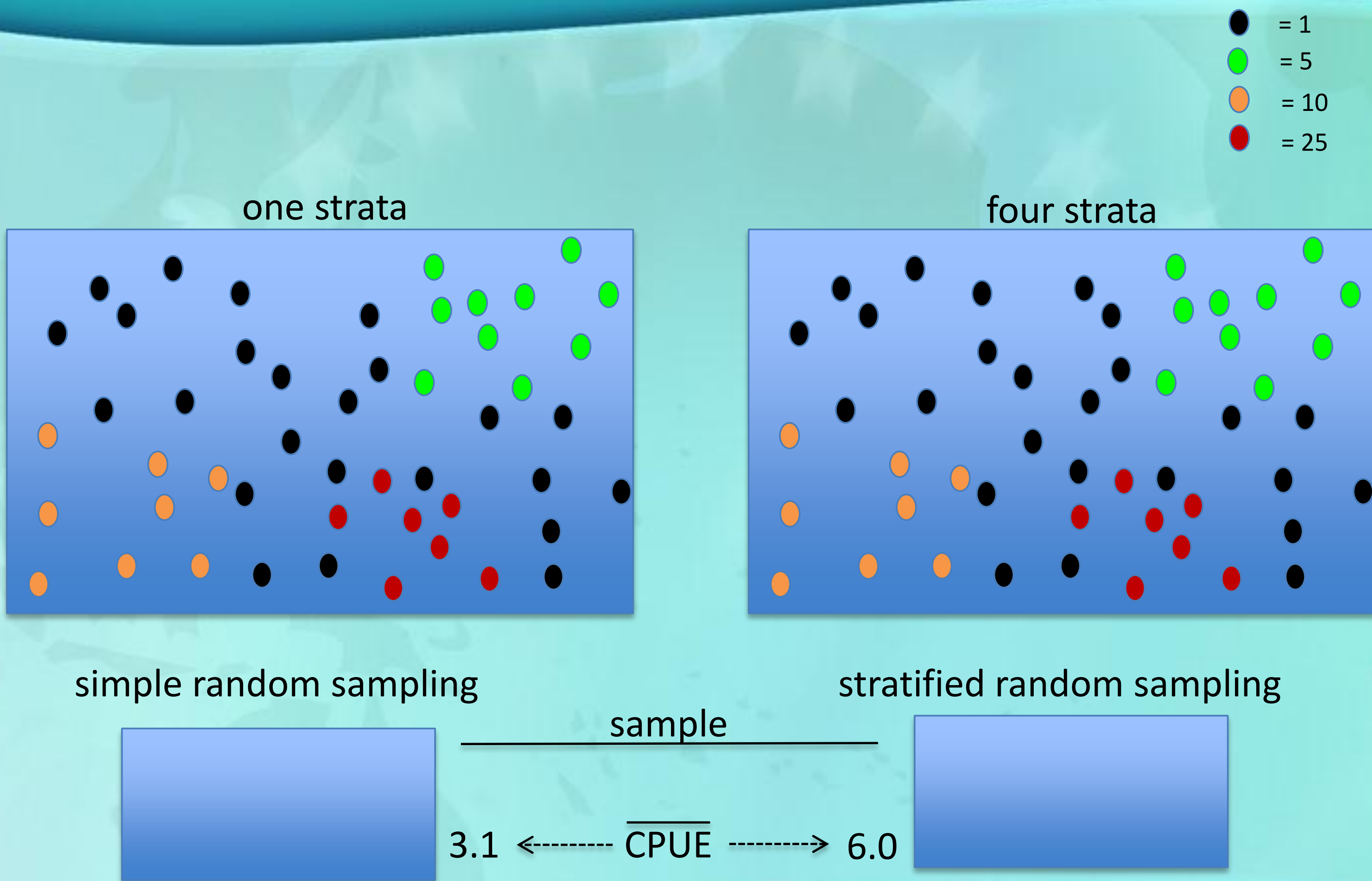
simple random sampling

stratified random sampling



sample



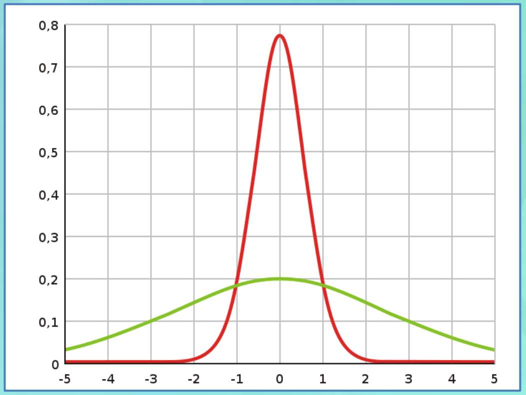

= 1
= 5
= 10
= 25
one strata
four strata
simple random sampling
stratified random sampling
sample
3.1 <--------- CPUE ---------> 6.0

SAMPLING STRATA

Think carefully about the main factors that can affect catch and effort variation - unique to your fishery

- Vessel size class: catchability, costs differ
- Gear: selectivity of fish of certain sizes (ages)
- Spatial: location of ports, landing sites
- Temporal: seasonality, monthly, weekly, time of day....
- Social/cultural: sacred times....

SOME SIMPLE STATISTICS

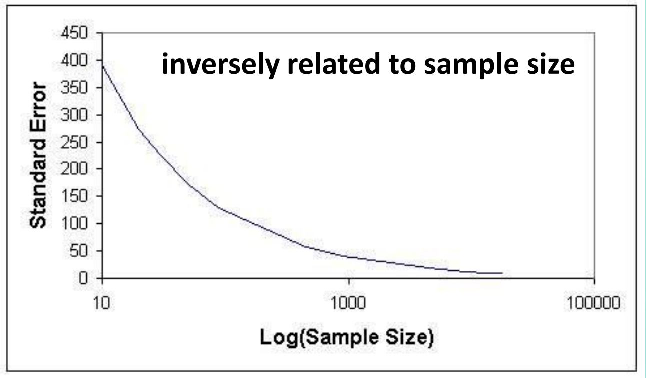


Samples are often summarized by taking an average.

Variance measures the overall spread of the data.

But how good is the estimated average itself?

Because samples were gathered randomly (probabilistically), the expected deviation in the calculated average can be computed – standard error.



Standard error provides the expected deviation between the sample average and the true population average (based on theoretically repeating your sampling procedure many times).

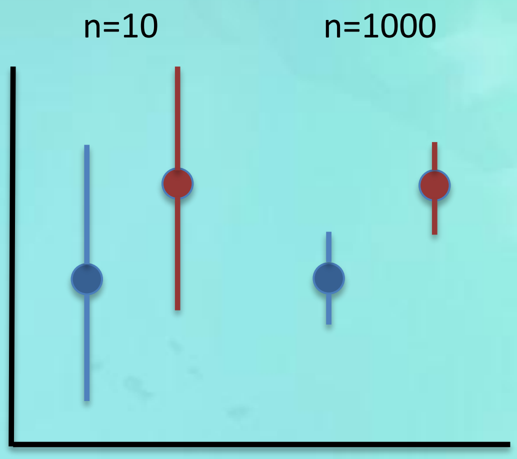
SOME SIMPLE STATISTICS

Statistical significance - a result that is not likely to have occurred randomly (or just by chance), but rather is likely to be attributable to a specific cause.

The fewer the samples, the less likely it is to have a <u>statistically</u> significant result.

Q: What does all this have to do with setting up a data collection programme?

A: Provides insight into the level of sampling intensity needed to facilitate a desired level of precision associated with resulting estimates.



SAMPLING INTENSITY

General - tradeoffs

- few samples provide little information (potential for bias, imprecision, and an inability to detect statistically significant changes)
- many samples (or a complete census) is usually cost prohibitive or impractical

“Rules of thumb”

- Samples: 30 samples per strata gives a ‘good’ average (not always feasible)
- Coverage: > 5%

It is often difficult to know the exact level of sampling intensity required without having a thorough knowledge of how variable the samples are expected to be.

Need to start somewhere and adapt as knowledge increases

STATISTICAL POWER TO DETECT TRENDS

Statistical power analyses can be conducted to find the sample size that would be required to be reasonably likely to detect an effect (e.g., statistically significant change in catch rates from year to year) of a given size.

It can also be used to calculate the minimum effect size (change in catch) that is likely to be detected using a given sample size.

Running statistical power analyses is possible to estimate necessary sample sizes but this requires assumptions about the inherent variability in the data (for each strata). Thus, the utility of these analyses can be limited in situations where data collection schemes are relatively new.

"I can see your error bars using google earth"

LOGSHEET COVERAGE AND 'RAISING' CATCH AND EFFORT

Logsheet coverage should be as high as feasible and (if not 100%) should come from a 'representative' group of fishers from each identified strata

$$\boldsymbol{Total\ catch} = \sum_{1}^{Strata(i)} \boldsymbol{Logsheet\ catch/trip}_{Strata(i)} * \boldsymbol{\#Trips}_{Strata(i)}$$

However, note that any bias (or mis-representation) in the logsheets when coverage is low is exacerbated when logsheets are raised to get the total catch and effort.

	Sample Design	Sample Size	Bias	Coverage	Raised Value
Country 1 -------->	Good	Moderate	2	80%	small bias
Country 2 -------->	Good	Moderate	2	10%	moderate bias
Country 3 -------->	Lacking	Low	7	5%	large bias

SUMMARY

Long-term data are necessary to tease out the likely cause of trending populations (e.g., due to seasonal or annual environmental fluctuations or due to fishing intensity).

Samples should be random and representative.

Account for important differences (random stratified sampling) to best estimate statistics (e.g., average catch/trip or # trips/day).

Sampling intensity should be as high as practically possible, but generally 20-30 sampling units per strata is good (< 5 does not give you much confidence in your average).

If your sampling design is sound, you can get away with less intense sampling (your overall estimate will be good b/c it 'incorporates' the major subtleties in the fishery and thus is representative).