The MED&BS data call assumptions (Statistical System)

ALWAYS consistent with the sampling design

Statistical Methods WS
under RDBFIS

Virtual January 2025

Huston …. or at least a misunderstanding

Mostly Med&BS data call and to a lesser degree RDBFIS does not hold DATA (sampling unit), it hold ESTIMATES (aggregated samples).

This is an important distinction, becau
Implementing the 'appropriate' raising
artefact of the design but also its use.

The same applies to the 'sampling unit' anchored in the sampling design. Its not that data has no restrictions, just that it has the least restriction. **Post Stratification**

You can always go forward but never back, consider these question:

What is the estimated selectivity of a métier?

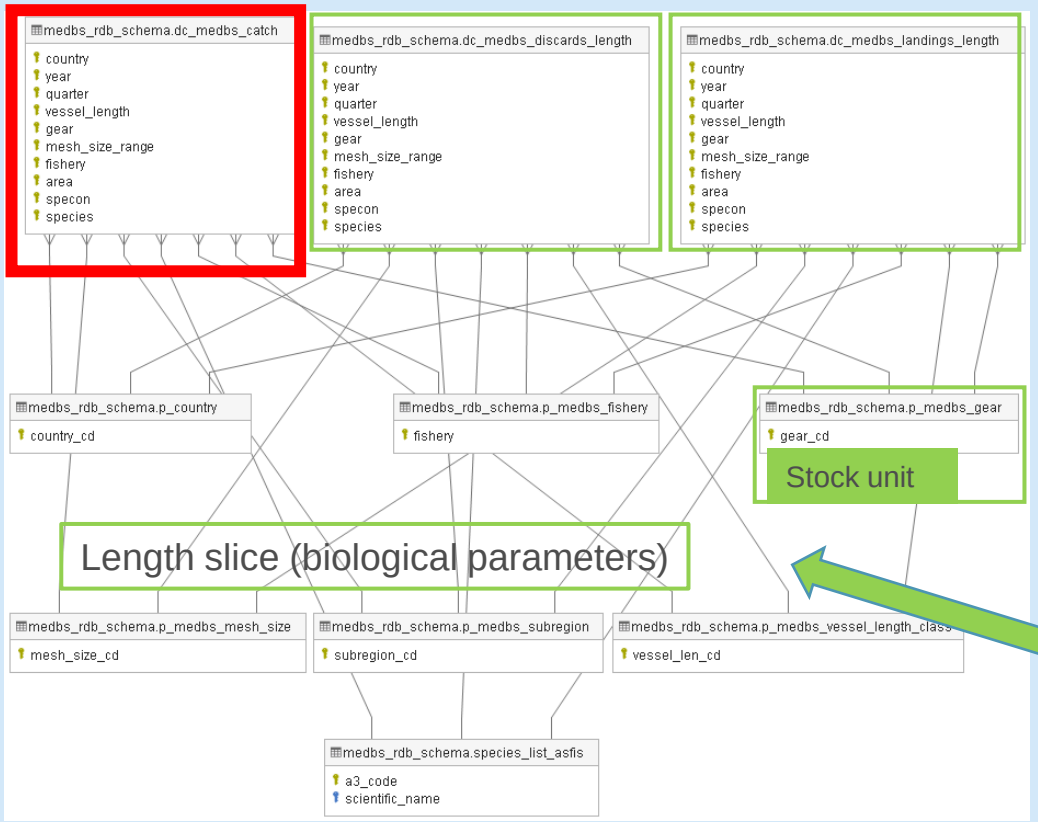
Are there differences in selectivity between me

But it's a 'mixed bag'. Give an inquisitive person enough rope and they will hang themselves. **Post Stratification**

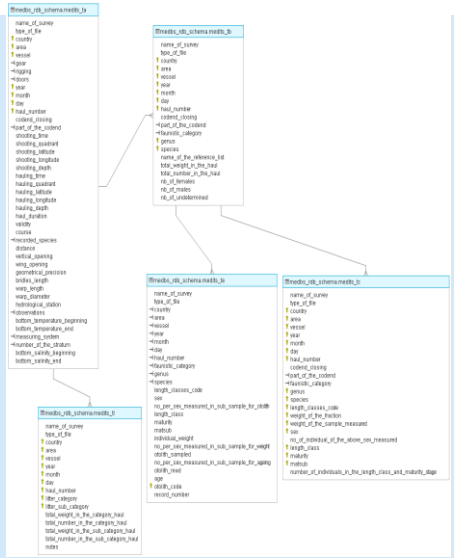
Is there really a statistical system for the MEB&BS data call? (Thanks Stefanos)

Acoustic Surveys

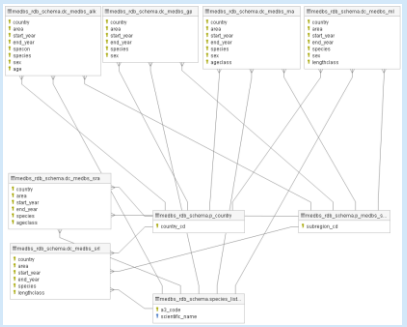
Catches



Demersal Surveys



Biological Parameters



DS data is data! EWG uses its own statistical system (stock unit), but largely similar

European Commission

Its all about ‘POPULATION’ and ‘REPRESENTATION’ and its nothing to do with American Elections

We want catch@age from nested design

To get there we need several raising factors and aggregations. In reality the process is one of scaling things ‘appropriately’ and the stock assessment is just another method of scaling in this process

HOPE

Maybe

NO

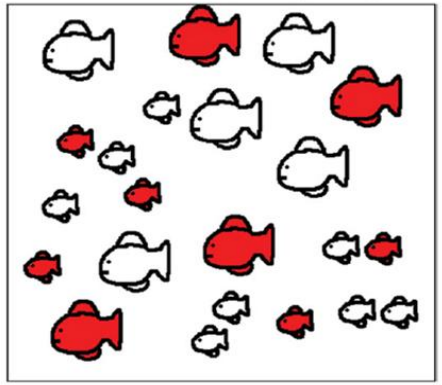
NO

Scaling catches to the stock by numbers-at-age. (Assessment)	Numbers-at-age in catch summed across all fleets	Numbers-at-age raised by weight to get numbers-at-age (by fleet)	Numbers-at-length raised by length class to get numbers-at-age (Age-length-key)
n@age	weight	weight	length

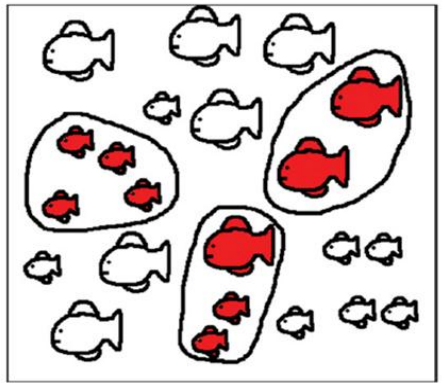
Why are they not representative of the parent population?
Do we fix it?

So…. what's the issue? Sample numbers, sort of

Random Sampling



Cluster Sampling



Nelson, G. A. (2014). Cluster sampling: a pervasive, yet little recognized survey design in fisheries research. *Transactions of the American Fisheries Society*, *143*(4), 926-938.

Who's fault is that? Are we not doing our job?

We can't sample the whole catch at the same time.

Samples are autocorrelated and the degree of autocorrelation is unknown.

Season, fleet availability, area …

This has impacts on the error distributions that are assumed in the stock assessment

- Stock assessments assume fully mixed populations. I don't know any cases that are.
- Different assessments make different assumptions (this is a reason for choosing a model?)
- Likelihoods / statistics / conclusions are conditional on the assumptions around the error structure. You tend to increase the type II error rate.

Conclusions! How do you decide to aggregate the stock data?

- Its purpose dependent. What do you want this data to be used / usable for?
- Are you deciding the data use? RDBES seems to think so. But are you sure what questions the commission wants answered now and into the future.
- How sure are you of randomness within the STRATA / PRIMARY KEYs
- Is there a RIGHT way to aggregate data (usually no), is there a best way given a particular model (almost certainly).

Keep in touch

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub – Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

@eu_science

European Commission

Thank you

European Commission

An Overview of sampling design and data use

We ‘understand’ the population dynamics in the stock

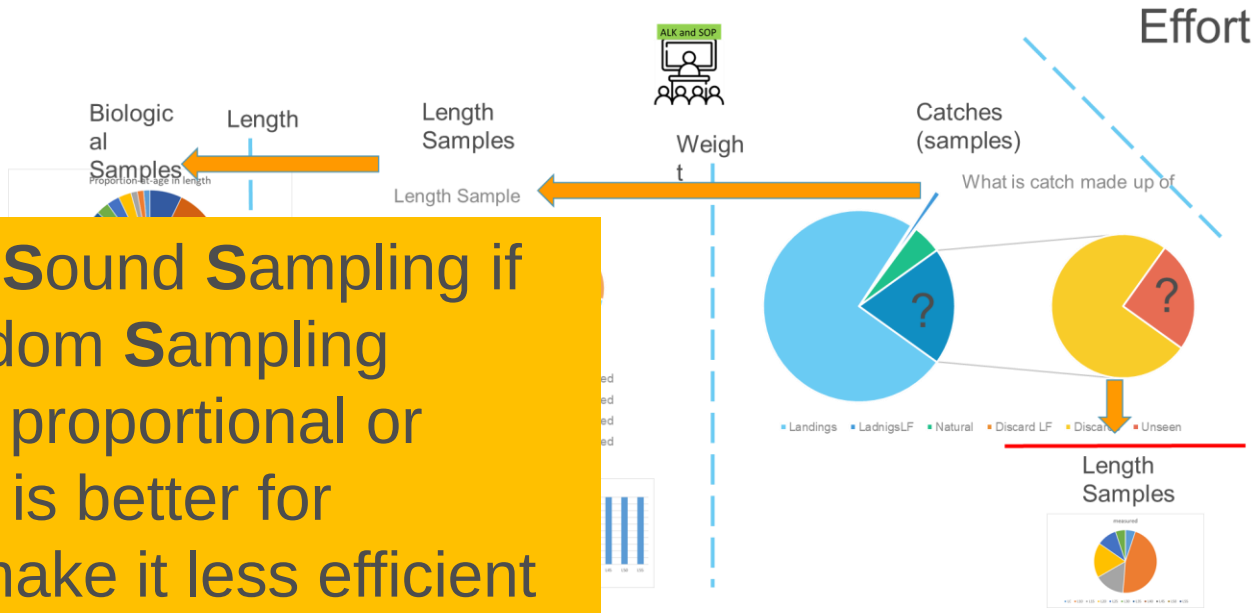
What are the processes and the pa

When we sample that population we get data.

How do the data SA inpu

- What is **S**tatistically **S**ound **S**ampling if not a **S**tratified **R**andom **S**ampling design??? It can be proportional or weighted. The latter is better for efficiency but may make it less efficient for other metrics and complicates post-stratification efforts or model based indicators.

+ Survey



al’ & ?

To have a hope in recovering the dynamics in our **stock** our aggregation has to reflect the dynamics. (It HAS to represent the **statistical populationS**)

European Commission

Variation is the sample deviation from the true value in the population that the sample comes from

Is varia

Is our sar
STANDAI

Represer

Is a RANI

Can we te

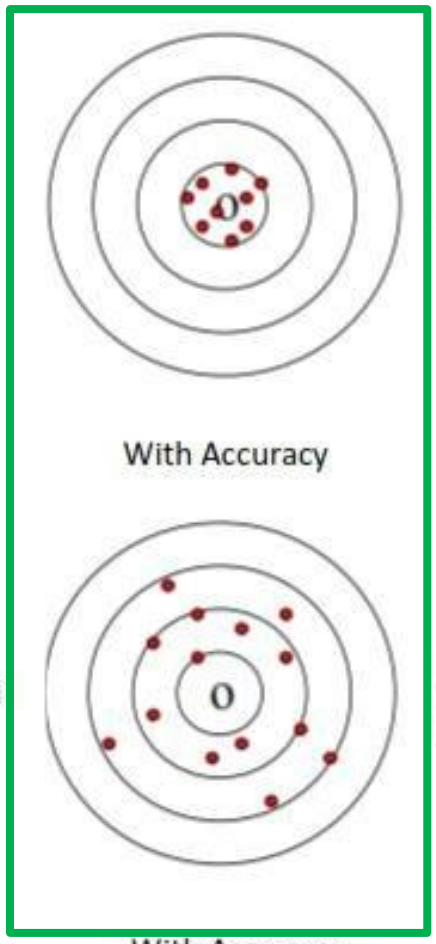
Rememb
of the **cer**

The medi
i.e. 50% a
central te

When all factors that affect the outcome are also random.
STRATUM

Represent
Sampl

What
sure
repre



How do we control variance?

How do we control bias?

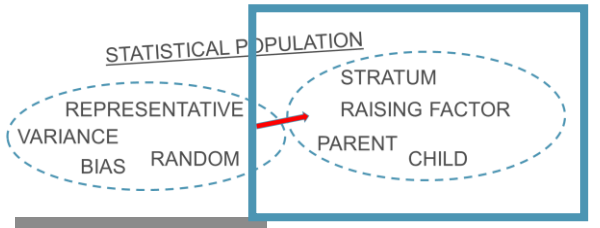
What is the interaction?

small group
perties of the
re population

WallStreetMojo

European Commission

Some more words from the **relational** database world. ”PARENT, CHILD, PRIMARY-KEY”



Data we need in the same table

PARENT CHILD

PARENT CHILD

Childless PARENT

HAULS

Survey Ha	Day	Stratum
1	2	A
2	2	A
3	3	B

CATCHES

Survey Ha	species	catch weig
1	MUT	5
1	HKE	2
1	ARA	7
3	ARA	5
3	ARS	3
3	HKE	20

LENGTH FREQUENCIES

Survey Ha	Species	Length	Number	Sample Weight
1	MUT	25	1	5
1	MUT	26	1	5
1	MUT	15	1	5
1	HKE	18	5	2
1	HKE	2	2	2
1	ARA	1	1	7
1	ARA	2	20	7
1	ARA	3	70	7
1	ARA	4	10	7
1	ARA	5	1	7
3	ARA	5	8	1
3	ARA	6	15	1
3	ARA	7	13	1
3	ARA	8	5	1
3	ARA	10	1	1
3	ARS	12	5	0.5
3	ARS	13	5	0.5
3	HKE	50	1	20
3	HKE	69	1	20
3	HKE	71	1	20
3	HKE	80	1	20
4	HKE	90	1	3

REG-KEY-1

REG-KEY-2

We use the STRATUM as the PRIMARY-KEY that LINKs the tables

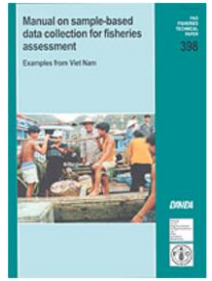
Each Child has exactly one corresponding parent, but a parent can have many children. Parentless children are not allowed in a database, **but sampling can have them and so can EXCEL!**

Parentless CHILD

Survey Hau	Day	Stratum			species	catch weight			Length	Number	Sample Weight	
1	2	A			MUT	5			25	1	5	
1	2	A			MUT	5			26	1	5	
1	2	A			MUT	5			15	1	5	

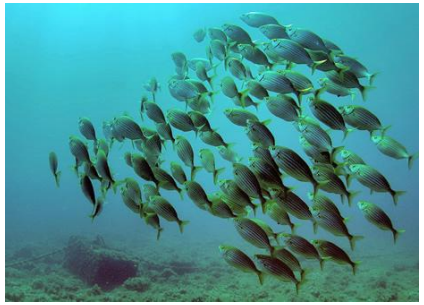
European Commission

How we generally sample fisheries (looks familiar?)



THE STRATUM
Primarily variance reduction

WE WANT
The stock



What other Effects do we model?

?

The effect

Total landings in weight

?

The "fleet", "quarter", "area"

Fleet catch in weight

?

The fishES
autocorrelation

?

Numbers at length

?

The fish disproportionate sampling

Length
Sex
Age
Weight
Maturity

The other …post stratify?

What inputs do we get from this?

Proportional with respect to variable of interest in relation to the population of interest.

European Commission

Sample Raising (principle)

is the process of accounting for the proportion of the “target metric” that we have more detailed information **from / for**.

For nested sampling designs (example: fleets) this implies the ratio of the same metric between the parent and child stratum multiplied by the desired metric from the child stratum missing from the parent stratum (fish)

Fleet caught 100t / Our LF sample weighed 100kg = 1000

Raising factor (RF)

Ca. 345000 fish in fleet catch = Sample contained 345 fish * 1000