fish5101fishsci Introduction to fish population dynamics

Gunnar Stefansson

19. desember 2016

Acknowledgements
MareFrame is a EC-funded RTD project which seeks to remove the barriers preventing more widespread use of the ecosystem-based approach to fisheries management.
`http://mareframe-fp7.org`
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no.613571.
`http://mareframe-fp7.org`
Háskóli Íslands
`http://www.hi.is/`

Efnisyfirlit

1 Introduction and overview

1.1 Introduction

- Introduction to fish population dynamics
- Different form of advice, data collection, etc.
- Available at http://www.tutor-web.net

1.1.1 Details

The documentation for this course is provided in several formats: tutorial notes (PDF), Tutor-web, and lecture slides. The PDF provides the lecture material while the Tutor-web page contains the quiz questions. Thus, it is my intent that you will read the PDF and then take the associated quizzes on Tutor-web. The lecture slides can be largely ignored.

Fish stocks are resources which are renewable if they are correctly handled. Obviously no large stock will be harmed if only a relatively few individuals are caught from the stock. Similarly it is clear that if all mature fish are caught before spawning, then there will be no recruitment. Thus it is essential to ensure that there is a sufficient amount of mature fish in the stock at each time. The problem is to find the appropriate middle ground where it is possible to obtain good catches for a long time.

In the following pages an attempt will be made to go through the main points which relate to the estimation of the size and productivity of fish and whale stocks. Examples will relate to models which have been used for the estimation of the size and productivity of stocks in the Economic Zone of Iceland.

1.2 Some possible textbooks

Primary references

- Haddon: Modelling and quantitative methods in fisheries. Chapman and Hall. (ISBN 1-58488-177-1)
- Terrance J. Quinn II and Richard B. Deriso: Quantitative fish dynamics. Oxford Univ. Press. (ISBN 0-19-507631-1)
- King: Fisheries biology, assessment and management. Fishing News Books. (ISBN 0-85238-223-5)

Alternative references abound.

1.2.1 Details

A large number of text books are available in fish population dynamics. The"bible" of fisheries is the book of Beverton and Holt, but it is somewhat dated. The primary alternative reading for this course are the books by Quinn and Deriso and by Haddon, which are the closest to the topic presented here. The book by King is also an excellent read.

A recent book by Hilborn and Walters' takes the reader in a slightly different direction, emphasizing Bayesian and risk analysis approaches.

Totally different approaches are also available, including the highly mathematical approaches of Murray as well as the exact opposite descriptive approach of Ross.

As far as this on-line course is concerned the tutorial notes provided on tutor-web.net, should be sufficient. However, it is always useful to consider alternative treatments of the topic.

Some of the more important books follow:

- Haddon: Modelling and quantitative methods in fisheries. Chapman and Hall. (ISBN 1-58488-177-1)
- Terrance J. Quinn II and Richard B. Deriso: Quantitative fish dynamics. Oxford Univ. Press. (ISBN 0-19-507631-1)
- King: Fisheries biology, assessment and management. Fishing News Books. (ISBN 0-85238-223-5)
- Beverton and Holt
- Cushing: Marine ecology and fisheries. Cambridge Univ. Press
- Diana: Biology and ecology of fishes. Biol. Sciences Press
- Elliot: Quantitative ecology and the brown trout. Oxford Univ. Press
- Getz and Haight: Population harvesting. Princeton Univ Press
- Hilborn and Walters: Quantitative fisheries stock assessment. Chapman and Hall
- Moss, Watson and Ollason: Animal population dynamics. Chapman and Hall
- Murray: Mathematical biology. Springer Verlag.
- Ross: Fisheries conservation and management. Prentice Hall
- Royama: Analytical population dynamics. Chapman and Hall
- Seber: The estimation of animal abundance. Edward Arnold
- Wood and Nisbet: Estimation of mortality rates in stage-structured populations. Springer-Verlag
- Wootton: Ecology of teleost fishes. Chapman and Hall

1.3 Terminology

Prefer "fish population dynamics and rational utilization"to "fish stock assessments".

1.3.1 Details

Although it is a common desire to want to know how many fish are in the sea and therefore use "fish stock assessment"it is a bit of a misnomer. The emphasis needs to be on how the stocks can be utilized or on how their numbers and biomass are likely to develop given specified conditions.

For this reason the preferred terminology is "fish population dynamics". The term "rational utilization” of fish stocks is also a much better term than "fish stock assessments” since it is rather useless to know the number of fish in the sea without knowing the potential yield of the stock.

Note 1.1. For the remainder of this course, the term "fish stock assessment"will be restricted to the actual estimation of stock size.

Remember that stock size is a relatively minor part and that an understanding of population dynamics is needed in order to figure out a rational utilization of the resource.

1.4 Motivation, assumptions, methods

Things to keep in mind when providing fisheries advice:

- Purpose: Short-term advice, general understanding, or long-term view on utilization?
- Accuracy: In terms of advice, general trend/knowledge, stock size, ???

1.4.1 Details

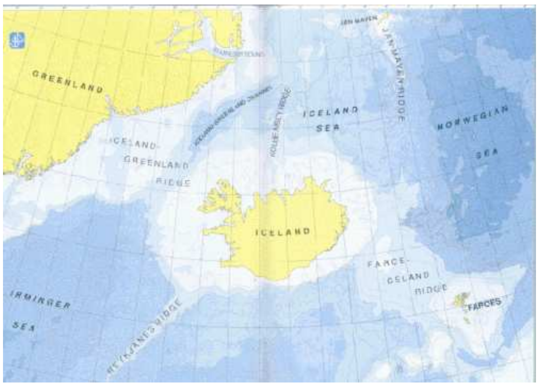
The handling of this topic is necessarily tainted by the fact that one of the main tasks of fishery scientists is advice on fishing. In this context, a clear distinction must be made between advice and management. Advice is intended for those in charge of fishing in order for the managers to be able to make decisions with the best available knowledge about the resource.

Fisheries advice comes in different shapes and sizes but most commonly attempts are made to answer the question, "How will a particular catch regime affect the future stock and yield potential?". Therefore advisory bodies usually first try to assess the size of the stock and then to predict the consequences of different catch levels.

Naturally a manager may ask different questions,"How can a stock be utilized in the long run” or "What will be the likely consequences of a particular action on catches and the development of the stock?". In some cases an adviser is asked to construct an entire management system. Such a system can be viewed as a computer program which inputs biological information and returns a quota, effort-days, or other quantified management measures. The design of such a system must take into account all available information about the stock. In addition, the system must be extensively tested, taking into account the fact that there is always considerable uncertainty in the estimate of stock size at any given point in time as well as in the initial biological information available at the onset.

Long-term utilization of resources is commonly skimmed over or largely omitted from course content as well as annual advice on the utilization of fish stocks. This issue is crucial, however, since it is possible to set yearly catch levels which will seem reasonable in the short-term but are catastrophic in the long run.

1.5 Typical stock areas, c.f. redfish



Typical redfish areas stretch across several economic zones.

1.5.1 Details

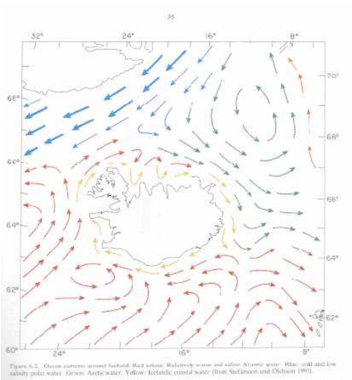
The primary focus of most scientific research is first to obtain an overall picture and later to fill in detail. Thus, one should always begin by trying to understand how a species behaves, then how a stock is likely to respond to things like fishing pressure, and finally, at a much later stage, worry about stock components.

Ignoring stock components can certainly lead to problems, but it is much more serious to go into such detail initially as it can render advice on the utilization of fish stocks useless.

1.5.2 Examples

Example 1.1. When determining fisheries advice for redfish (*Sebastes marinus* and *S. mentalla* and others) in the central north Atlantic, one initially considers large areas, such as the one in the figure, East-Greenland-Iceland-Faroe Islands.

1.6 The physical environment



1.6.1 Details

It is of considerable importance to have some knowledge of the driving forces in the ecosystem. This includes the basic biology of the species including spawning grounds, larval drift routes, growth parameters, food supply, general migration routes, etc.

Of fundamental importance is a knowledge of what drives variation in these various parameters. For example, it is well known that physical conditions in waters around Iceland have considerable influence on the growth of many species as well as on recruitment variation.

Such issues need to be kept in mind when utilization is considered.

1.7 Advice with ample data but no strategy

Cod in Icelandic waters 1993-94
Background:

- Biomass in 1993 estimated 630 000 t (640) and SSB (spawning stock biomass) at 210 (230).
- Mean weight increasing.
- Maturity at age was high in 1993, but lower than in 1992.
- Average recruitment between 1985 and 1992 was only 128 million 3 year olds. The 1991 yearclass had an average recruitment of 73 million 3 year olds. The long-term average is 220 million.
- Catches in 1993 were predicted at 230 000 t - 31% over advised level.
- Fishing mortality was far above F0.1 (a fishing mortality in which on 10% of the marginal yield-per-recruit is harvested).
- SSB in 1994 only about 200 000 t (close to historical minimum)

Prospects:

- Increased likelihood of continued poor recruitment
- Biomass in 1994 at 610 000 t (historical minimum)

Proposal:

- Proposed catch limit 1993/94 150 000 t.

Later:

- 1994 adoption of formal catch control law
- controls and assessments failed
- controls and assessments were made more restrictive in 2007

1.7.1 Details

The type of advice give varies depending on the seriousness of the state of the fish stock as well as the general framework within which the advice is given. Thus, in a situation where there is no system to regulate catches, there is little purpose in providing recommendations on reducing catches. There might, however, be reasons to provide advice on the likely catch in the coming years, given e.g. a continuation of the present effort.

When there is a system to regulate catches and there is a clear case of overexploitation the advisers must give clear advice on catch reductions which will improve the situation.

In cases where there is ample data it should be possible to formally estimate the status of the stock and give future projections. It is then also possible to indicate which catch levels are likely to lead to an improvement in stock status.

1.7.2 Examples

Example 1.2. Cod in Icelandic waters had reached a severely depleted state by the early 1990s. At that time, it was clear that strong advice needed to be given.

In 1994 a formal catch control law was adopted which set limits on the number of fish harvested. Initially this appeared to work but later on (ca 1997) controls and assessments failed.

The catch control law (CCL) was revised several times, in particular a more restrictive CCL was implemented in 2007.

1.8 Advice with poor data and no strategy

- Catch in 1992 was 94 000 t.
- CPUE (catch per unit effort) surveys indicated a 50% reduction in 5 years
- Catch of redfish has been declining
- Effort should be reduced and total allowable catch (TAC) limited to 80 000 t.

1.8.1 Details

In many cases there is very little information available on the stock size. There may, for example, only be length distribution data available which provides very little information on the age composition or recruitment.

Even in data-poor situations it is the responsibility of the fishery scientist to provide basic information on how the stocks can be utilized in a reasonable manner.

If there is information about aggregate catches and some index of total abundance, this likely provides the means to make sensible statements about the state of the resource and likely directions.

1.8.2 Examples

Example 1.3. Even at the end of the 20th century there was very little biological information on the stocks of redfish in the North-East Atlantic and advice reflected this lack of information. Specifically, around 1990 data on species composition around Iceland did not exist.

The approach taken was therefore to provide advice which used the available data, which was basically total landings and a survey index of total abundance. Given that the CPUE in the available groundfish survey had declined by 50% in the previous 5 years, it was clear that a continuation of such a decline would not constitute a sustainable fishery. Hence the advice given was intended to halt the decline.

1.8.3 Slide Reference

MRI 1995 State of Marine Stocks in Icelandic waters 1994/95. Prospects for the quota year 1995/96.

1.9 Advice in the presence of a strategy

Herring in Icelandic waters

- Agreed total allowable catch (TAC) was 20-25, which was above recommendations in past years
- Yearclasses did not last as long as predicted
- Large yearclasses: 1983, 1988, and 1989 sustained the fishery
- SSB around 400 thousand t.
- Increased recruitment with increasing SSB
- TAC recommendation at F0.1 - 90 000 t. (implicit CCL)

1.9.1 Details

From the point of view of an adviser on fisheries, the ideal situation is when a formal harvest control rule exists, including a catch control law (CCL). The CCL is simply a formula which specifies what the annual catch should be, given the available data.

1.9.2 Examples

Example 1.4. Herring in Icelandic waters has long been harvested with a target fishing mortality of $F_{0.1}$. This implies that the fishery scientist simply needa to evaluate the catches corresponding to this level and that the total allowable catch is the "advice".

1.10 Certification

Current trends are towards "certification"

In all cases this means:

- Harvest Control Rules
- MSY (Maximum sustained yield) definitions
- Application of the precautionary approach
- etc

1.10.1 Details

International markets increasingly insist that fisheries are "certified". This implies that the fishery needs to be able to demonstrate that it is biologically sustainable.

Requirements for certification are the same as FAO's Code of Conduct and other international agreements and definitions on what constitutes a sustainable fishery.

This course deals with these definitions.

1.11 Background information - Stocks, fleets etc

Stock units etc.

Population structure

Fisheries (meshes etc.)

Landings (total census, or sample)

Biological measurements (coverage, random, etc.)

Research vessels

Fishing vessels can be of all shapes and sizes.

1.11.1 Details

When considering the population dynamics of a fish stock, there are a number of biological concerns which need to be considered. Although not all of these are crucial at all times, in fact usually it is possible to get a general idea of stock development without consideration of all these issues, they need to be addressed eventually. Thus, sooner or later one needs fairly complete information on the stock units and other aspects of the population structure.

Of greater importance is the fleet composition and the conduct of the fisheries. Each fleet or gear tends to target a certain component of the stock, such as spawning fish, large fish, or small fish.

Sampling schemes for the purpose of understanding the dynamics or assessing the stock size need to take these issues into account. For example, sampling is often organized so as to adequately represent each fishery and/or landing site.

Finally, research surveys are frequently designed to cover the entire stock or as much of it as is feasible. Research surveys tend to be the most expensive portion of applied fishery science around the world. Research vessels are designed and built solely for the purpose of conducting surveys, which in most cases means monitoring of fish stocks.

1.12 Stock units and structure

Stock units:

- Need to know stock identity
- If many stocks or a substock, then ...

Population structure and biology:

- Age composition/Length distribution
- Growth
- Maturation

1.12.1 Details

Biological details eventually need to be considered not only for the purpose of sampling design but also in order to understand the population dynamics in more detail.

Examples include the stock units: If there are several substocks, then in principle it may be important to know whether a substock is being overfished in a small area. This might not be seen in a total stock estimate or in a global measure of abundance from a data set using the entire area.

Similarly, the maturation process is a function of many variables and for some species the onset of maturation has a major impact on subsequent growth. In this case it is important to know whether the maturation process is affected by fishing or other human activities.

In spite of all such complications, it must be kept in mind at all times that the important aspect of both stock assessments and fish stock utilization schemes is to have an overview. Many issues such as whether there are several spawning locations or the food supply is variable is often dwarfed by simpler facts such as a serious overfishing situation.

References ISBN: 1584881771
ISBN: 0195076311
ISBN: 0852382235
Beverton and Holt
Cushing: Marine ecology and fisheries. Cambridge Univ. Press
Diana: Biology and ecology of fishes. Biol. Sciences Press
Elliot: Quantitative ecology and the brown trout. Oxford Univ. Press
Getz and Haight: Population harvesting. Princeton Univ Press
Hilborn and Walters: Quantitative fisheries stock assessment. Chapman and Hall
Moss, Watson and Ollason: Animal population dynamics. Chapman and Hall
Murray: Mathematical biology. Springer Verlag.

Ross: Fisheries conservation and management. Prentice Hall
Royama: Analytical population dynamics. Chapman and Hall
Seber: The estimation of animal abundance. Edward Arnold
Wood and Nisbet: Estimation of mortality rates in stage-structured populations. Springer-Verlag
Wootton: Ecology of teleost fishes. Chapman and Hall

2 Collecting data for direct monitoring

2.1 Monitoring a resource

To understand the ecosystem you need to monitor it
$\Downarrow$
To monitor it you need measurements
$\Downarrow$
The preferred measurements are catches in numbers at age and survey index. Otherwise use dynamic age-structured production models etc. Can use only a survey index, only CPUE data etc.
$\Downarrow$
Must have consistent and annual sampling
$\Downarrow$
Data needs to be stored in a consistent format

2.1.1 Details

It is impossible to draw sensible conclusions without data. Using simple statements from textbooks or general theory tends not to be useful and is very often misleading.

In particular, a prerequisite for understanding the behavior of an ecosystem is to monitor it. For this purpose routine measurements are needed.

In this section a short description will be given of measurements commonly made for the purpose of monitoring fish stocks.

The simplest data for monitoring the health of a resources is in the form of total counts or weight of the catch taken in some standardized fashion. Thus, one can record the amount caught on average each year through the use of a specific gear in a number of locations. This provides an index - a measure of how the stock changes in time and will be considered in more detail later.

The best possible data for stock assessments are catches, with information on age and length composition along with survey indices (preferably also with age and length information). The simplest methods for such data are called VPA-based methods and are extensions of the original Virtual Population Analysis of Gulland (1965). More recent statistical methods are preferred, and it is through these methods that it can be verified how important the various data sources are.

If catches cannot be aged, production models are commonly used. Such models only need information on total catches in tonnes along with a survey index of abundance of commercial catch per unit effort. In order to estimate abundance and productivity it is necessary to have measurements which relate to the stock size in some way. It is rarely possible to estimate the abundance of a marine species directly, although exceptions exist such as the measurements obtained with acoustic surveys of pelagic fish and counts of whales. Even in these cases, however, there tend to be unresolved issues which imply that these measurements may be best treated as indices of abundance rather than as direct counts.

The most common situation is therefore that measurements are made of quantities which are only indirectly related to the stock size. The mathematical models must thus connect the measurements to the stock size and statistical techniques are used to estimate unknown parameters in such models.

2.2 Catches (landings): Notation

Notation for landings data:
Y =catch (yield) in tonnes (or 000 t)
y = year (integers, 1985, 1986 etc.)

For example,
Y_{1986} or Y_{85}
Y_y =total landings in year y

2.2.1 Details

In order to estimate the amount caught or landed in a given year, some sampling scheme is needed. This sampling scheme needs to take into account seasonal differences in the fishing process, landing sites and so on.

Definition 2.1. Notation for landings data:
Y =catch (yield) in tonnes (or 000 t)
y = year (integers, 1985, 1986 etc.)

For example,
Y_{1986} or Y_{85}
Y_y =total landings in year y

The catch figures are a very important basis to understand the effects of fishing on the stock. These numbers are usually a key component in the estimation of the size and development of fish stocks. This is not always the case, however, and there exist direct measurements of stock size. Such methods will not depend on the existence of catch data.

Catches or landings data alone are never sufficient information about stock size. For example, landings may be at a low level for economic reasons even though the stock is in good shape. Similarly, catches can be maintained at a high level right up to a stock collapse for many pelagic stocks.

2.3 Linking data to advice

Short-term catch depends on the

- type of gear, timing, and form of targeting
- spawning vs juvenile components
- large vs small fish
- mesh size or other regulation

Longer-term effect on stock depends on

- typical effort expended
- type of gear/mesh size

2.3.1 Details

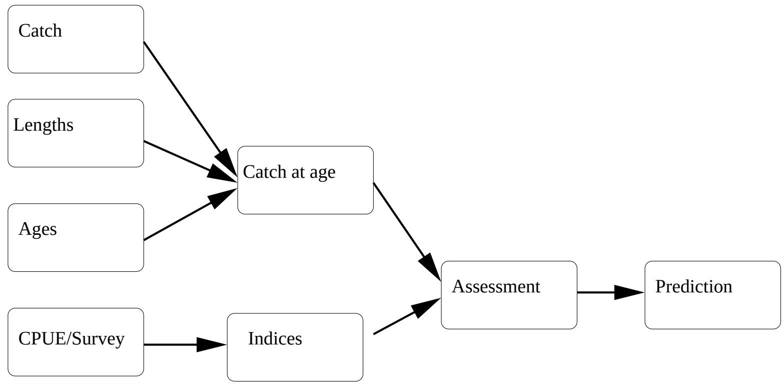
It is quite important to set up sampling and estimation strategies for catches that provide an overview of catches by gear, fishing area, and season. Ideally one should also have some idea of how the catches are separated into spawning and juvenile components, though this is sometimes either a consequence of location or estimated from biological data sampled from the catches.

Of course it is highly desirable to obtain reliable data in order to draw reliable inference. However, data may, in some case, not be particularly reliable, and it may be necessary to make some statement about potential yield or consequences of fishing from this sort of data. In such instances attempts are made to investigate how sensitive results may be in terms of changes in assumptions of variability in the input data. Thus, the approach may differ somewhat from the formal statistical methodology, where one will not draw conclusions unless there is a "significant" relationship.

In fishery science one may need to give appropriate recommendations, sometimes in situations where the data are very poor indeed. It follows that one should not think about estimation of catches or biological sampling of catches as a purely statistical issue. Rather, it is a part of an entire advisory process which can handle data at very different levels of accuracy.

The types of advice to be given can be from rough guidelines on long-term yield to short-term predictions of appropriate catch levels. These different questions imply varying levels of data requirements.

2.4 Traditional fisheries advice - data flow chart



Traditional data flow chart of fisheries data. One or more data boxes may be missing.

2.4.1 Details

The approaches to fish stock assessments vary considerably depending on data availability and the purpose behind the assessment. In the straight-forward case of annual assessments with annually collected age readings of high quality, the process is as follows:

Each year, data is collected by estimating the amount landed, along with biological sampling from the fishery and surveys.

Traditionally, the biological samples from the fisheries are used to compute the catches in numbers at age (along with various other important indicators such as mean weight at age).

The survey (or catch per unit effort from the fleet) data provides information on the relative stock status, in the form of indices of abundance. Similarly, underwater visual surveys provide indices of abundance which can in principle be used to monitor trends in abundance.

Acoustic measurements and sighting surveys (counts) of whales are of a different nature from the most common (indirect) measurements of marine animals when used as absolute abundance measures. Such direct measurements, which in principle provide absolute counts rather then indices, will not be covered in any detail, but they can also be used as indices, i.e. in the same manner as any other survey data.

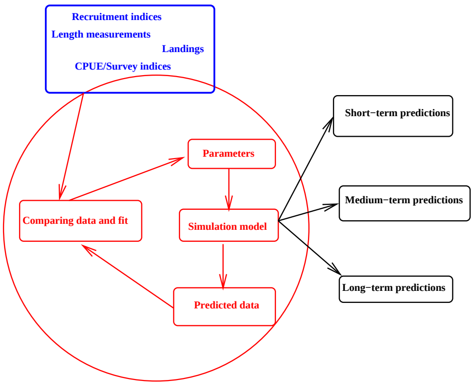
The same applies to tag-return data. Such data can be extremely important, but have mainly been used to obtain information on migration between regions.

Many of the above monitoring methods are commonly scaled up to give an absolute measure of abundance. Thus, in principle one can scale results from an underwater visual survey to obtain an estimate of total abundance. It will be seen, however, that this is a very dangerous procedure and will almost never be recommended.

Most commonly the assessment process consists of annual data collection followed by data summaries for each annual data set. The data summaries are subsequently used with mathematical models of the population dynamics to estimate the state of the stock and provide advice on its utilization.

In many formal statistical assessment models the procedure commonly deviates a bit from the above in that the data tend to be used in less of a summary form. Thus, length distributions are typically only used in simple summary form, not raised to the catches. Each such data set is simply used as a different piece of information in a so-called likelihood function.

2.5 Fisheries advice using generic data sets



Data collection, stock assessment and predictions. Blue denotes data collection, red the assessment phase, and black the prediction phase.

2.5.1 Details

Note that when only basic data sets exist, the need for formal modelling increases. In recent decades the trend has been to use formal statistical approaches. Thus, for any combination of data sets one first writes down an internal model describing the population dynamics, predicts the data, and subsequently finds the values of unknown parameters which best predict the data.

Note that internal models may include several processes which can not be measured. For example it is quite common to explicitly model age even if there is no age data. This "internal model"is usually a forward computation or "simulation model".

In order to make specific statements it is usually important to have actual data relevant to the question at hand:

- For the effect of fishing you need data on catches
- For stock trends, stock size/trend data should be used
- For prediction statements it is best to have recruitment data

One may be able to evaluate historical trends without any recruitment information and one may be able to evaluate the current exploitation rate without data on catches or effort. However, these data will be needed if one is to do reasonable predictions or provide management advice on the effect of reductions in catches or effort.

The full procedure can be summarized as follows:

1. Write a (mathematical) description of stock dynamics
2. Combine with available data for an assessment
3. Add short-term prediction (with uncertainty) of actions
4. Evaluate medium-term effect of actions

2.6 Catches or landings

- Knowledge of landings is essential in most cases
- Census is the "accepted practice"
- Can also (randomly) sample from chosen vessels (cheaper and sometimes better)
- Difficult to do any monitoring or control without knowledge of catches

2.6.1 Details

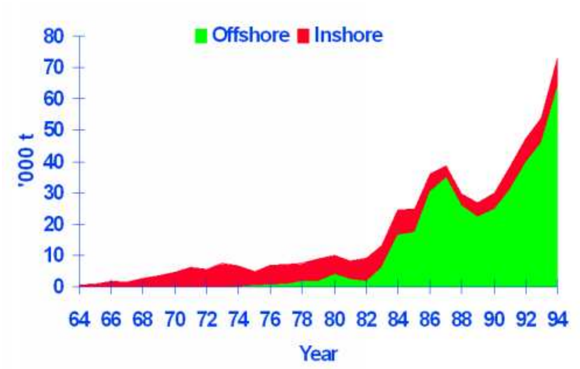
Note 2.1. The terms "catches", "landings"and "yield"will be used interchangeably in this course. Naturally, catches refer in principle to the actual amount caught whereas landings refer only to the amount brought to shore, the difference being discards.

In cases where there may be serious discards, these need to be estimated. In cases where the discards are unknown, deviations from assumptions can be investigated by exactly the same methods and are used to investigate the effects of variable (and unknown) natural mortality.

Note 2.2. Finally, "yield"in some circles refers to economic yield, but throughout this course it will simply be a synonym with catches and landings.

The catch figures alone do not say much about trends in the stock or its yield potential. However, it is clear that if a stock has given stable and good catches for hundreds of years, then it is likely to continue to do so. On the other hand, it is not possible in the short run to see whether the stock is being reduced by increased effort or technological advances. In order to separate these effects, it is imperative to obtain further information and it is important to have data on the age composition of the stock along with the effort or catch per unit effort along with survey information.

2.7 More on catches



Catches of pandalid shrimp in Icelandic waters, separated into offshore and near-shore (inshore) catches.

2.7.1 Details

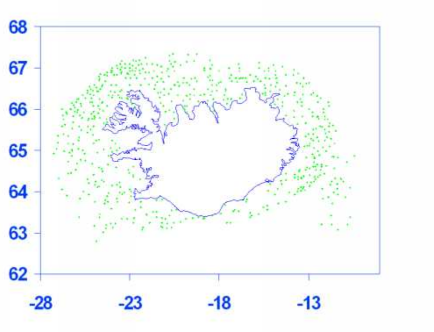
This example clearly shows how important it is to have appropriate data collection schemes. In this case the stocks are very different and it is important to know how the fleet changes behavior.

The fishery on the near-shore shrimp stock started in the late 1960s and provided steady but rather low catches throughout the remainder of the century. One reason for the stability is that the presented data is the sum of a number of smaller regions, each of which is quite volatile.

The offshore shrimp fishery is of a completely different nature, being based on larger vessels capable of fishing in all seasons and most weather regimes.

2.8 Research surveys

- Research vessels or other CONSISTENT and ANNUAL abundance measurements
- Usually designed to give time series of abundance, to be related to population trend
- Typically selects more age groups than commercial catches - e.g. catch younger fish than fishermen
- Need to cover stock distribution
- Want to be able to predict future recruitment
- Design is important! Define fishing gear and timing based on biology of the species
- Prefer age information but can use length-based or aggregated



2.8.1 Details

Note 2.3. The use of a standardized bottom trawl on a research cruise is referred to as a groundfish survey or a bottom-trawl survey. The trawl is used in exactly the same way every year and measurements are made of its contents from each tow.

The bottom trawl is not a sacred gear . It is possible to use other gear and obtain similar indices, but it can be difficult to, for example, to use a longline because its catchability depends on food supply while gillnets will select for specific size groups of fish.

Alternative methods include Underwater Visual Surveys (UVS), usually conducted with divers but potentially also using Autonomous Underwater Vehicles (AUVs). In principle these give absolute stock estimates but several sources of bias and variability exist and these must be considered for these as well as any other surveys.

Whatever method is used, standardization and consistency are essential. A survey needs to be conducted in exactly the same fashion for many years in order to obtain a view of the dynamics of the stock.

A survey which is only conducted once is useless as there is nothing to compare with. Two years of data will not say very much either since the variability in the survey tends to be high and can only be estimated or seen after more years have passed.

If the fishing gear used in a survey is changed from one year to the next, then that will typically render the survey useless as well. At best some estimate in the change in efficiency of the survey gear can be made but these estimates tend to be quite inaccurate.

In some cases a measure of total abundance is obtained by multiplying the average amount caught by a number which should indicate the catchability of the fishing gear. This practice has been largely abolished since it has led to many serious errors in real situations.

In addition to the above issues, surveys are a useful tool to provide biological data sets in a fishery-independent manner, as seen in following sections.

2.9 Survey indices

Bottom-trawl surveys: Obtain standardized catch-per-unit effort

Abundance indices: Should reflect trends in stock size

An Icelandic research vessel

2.9.1 Details

Note 2.4. Abundance indices simply refer to some numbers which are closely related to stock size.

One would expect that the catch per towing hour in a groundfish survey using a bottom trawl should represent a proportion of the total number of the fish in the sea, at least for those fish species which normally stay close to the bottom.

The catch in each tow during a survey will be very highly variable. Thus, it is important to take a large number of tows in order to reduce the variance in the resulting survey index. The statistical issues to be dealt with in this context are quite well known and the primary issue (apart from standardization) is always to obtain the largest possible number of tows.

Increasing various measurement types from each tow does not help the accuracy of the index - only increasing the total number of tows can reduce the variability. In some cases modelling approaches have been used to reduce the variability somewhat, but this will tend to be minor compared to the overall sampling error.

2.10 Data storage

- Must use consistent data storage "rectangular tables"
- Standardized for all species
- On an institutional level not the individual researcher level
- Easy extraction of all years into one file
- Easy combination e.g. of length and abundance information
- Formal data base (e.g. MySQL or PostgreSQL - both free)
- Not Excel, not Access

2.10.1 Details

Some of the principles involved when setting up a data storage system include:

- Must use consistent data storage
- Standardized for all species
- On an institutional level rather than individual researcher level
- Easy extraction of all years into one file
- Easy combination e.g. of length and abundance information
- Formal data base
- Not Excel or Access

It is amazingly common that entire institutes may have no formal storage system. This is all the more surprising in light of a few facts:

- Data storage systems are available for free
- If data are to be useful they need to be accessible
- Accessible means centrally accessible, not just on one laptop computer
- Standardization of the data sets for different years is the only way to sensibly do computation
- etc. etc. etc.

2.11 Basic design issues

Common design criteria include:

- The distance between $\bar{x}$ and μ should be no greater than δ,
- The confidence interval width should be no more than δ.
- The standard error of $\bar{x}$ should be no more than δ.

Fisheries data rarely satisfy the assumptions!

2.11.1 Details

Under simple random sampling from a Gaussian distribution with known variance the confidence interval has width $2z_{1-\alpha/2}\sigma/\sqrt{n}$ and it is not hard to see that to get the width of this below δ requires $n \geq (2z_{1-\alpha/2}\sigma/\delta)^2$. This is detailed somewhat below.

Note that **fisheries data do not satisfy these assumptions**! For that reason alternative approaches need to be considered. The methods below are a good start, however.

2.11.2 Handout

If simple random sampling applies, n data points come from a Gaussian distribution, and the variance (σ^2) is known, a confidence interval is obtained with

$$\bar{x} \pm z_{1-\alpha/2}\sigma/\sqrt{n}$$

The width of the confidence interval is

$$2z_{1-\alpha/2}\sigma/\sqrt{n}$$

Commonly, $\alpha = 0.05$ is used and then $z_{1-\alpha/2} = 1.96$.

The standard error of the mean is $\sigma/\sqrt{n}$.

Some design criteria include

- The distance between $\bar{x}$ and μ should be no greater than δ,
- The confidence interval width should be no more than δ.
- The standard error of $\bar{x}$ should be no more than δ.
- The coefficient of variation (of $\bar{x}$, in %) should be no more than δ.

Confidence interval width criterion:

$$\begin{aligned} & 2z_{1-\alpha/2}\sigma/\sqrt{n} \leq \delta \\ \Rightarrow \quad & n \geq (2z_{1-\alpha/2}\sigma/\delta)^2 \end{aligned}$$

Distance between $\bar{x}$ and μ criterion:

$$\begin{aligned} & z_{1-\alpha/2}\sigma/\sqrt{n} \leq \delta \\ \Rightarrow \quad & n \geq (z_{1-\alpha/2}\sigma/\delta)^2 \end{aligned}$$

Standard error of $\bar{x}$ criterion:

$$\begin{aligned} & \sigma/\sqrt{n} \leq \delta \\ \Rightarrow \quad & n \geq (\sigma/\delta)^2 \end{aligned}$$

CV criterion

$$\begin{aligned} & 100 \cdot (\sigma/\sqrt{n})/\mu \leq \delta \\ \Rightarrow \quad & n \geq (100 \cdot \sigma/(\mu\delta))^2 \end{aligned}$$

References

3 Use and design of biological samples

3.1 Biological measurements from catches

When collecting biological measurements from catches one needs to consider...

- Coverage: Needs to be representative!
- Randomization: Must not select for large fish, for example!

MUST HAVE A DECENT SAMPLING SCHEME

If not: Need totally different techniques for control, but theory still applies

Ideally: For each individual in the sample one would gather the following pieces of information:

- Length
- Weight
- Sex
- Maturity stage
- Age

3.1.1 Details

The fundamental measurements taken from biological samples of catches are length measurements and age determinations.

Since length measurements are more cost-effective than aging a fish, more length samples are taken than age samples. For this reason attempts are made to record as many age-specific attributes as possible. For example, fish are commonly weighed, their sex organs are weighed, their fat content is recorded, as well as other factor which are relevant to the understanding of the stock in question.

3.2 Length measurements

Length measurements can be obtained at a minimal cost from most fish stocks.

Usually, lengths are measured on a measuring board and simple counts are tallied as the measurements progress.

A better strategy is to simply record individual measurements. In this case one commonly also records sex, maturity stage, weight, etc.

Try to take random samples!!

3.2.1 Details

Length measurements can be obtained at a minimal cost from most fish stocks. Usually, the lengths of fish are measured on a measuring board and simple counts are tallied as the measurements progress.

```
cm
12
13 |
14 |
15 ||
16 |||
17 |
18
19 ||
20
```

For this reason it is possible to obtain very large numbers of length measurements at a low cost.

However, the length measurements do not provide an infinite amount of information. For example, although it is clear that increasing the number of length measurements will provide a very clear picture of the length distribution of the population, it does not follow that this increases knowledge of the age composition, growth, or other factors.

Similarly, care must be taken in how the fish are sampled for length measurements, as with any data collection. For example, measuring a large number of fish from a single tow will provide little information about other tows, even adjacent ones.

Sometimes fish are pre-sorted into size groups. This requires special attention.

3.2.2 Examples

Example 3.1. Sample data sets can be input into R using

```
dat<-read.table("http://www.hi.is/~gunnar/kennsla/alsm/data/set103.
   dat",header=T)
```

A length distribution can be obtained with

```
table(dat$le)
```

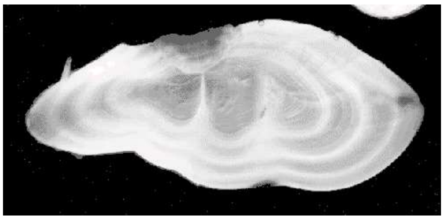
Note, however, that the zero-frequency lengths are missing in the resulting table. In order to obtain a table with those, one needs to set up an initial length distribution with all the length classes but zero frequency. The nonzero counts are subsequently inserted in the appropriate places.

```
ldst<-table(dat$le)
full<-rep(0,120)
names(full)<-1:120
lens<-names(ldst)
full[lens]<-ldst
```

Note that in this R example, the length distribution is indexed by the names of the cells.

3.3 Age readings

If there are age markers (rings), then one can sample randomly from catches to obtain the proportion of fish in each age group. To accomplish this a random sample is needed with good coverage of space and time for each gear type.



3.3.1 Details

Even in species where we may not be able to observe cohorts, they form a fundamental concept since every fish ages each year it survives. The age provides a natural basic time unit and is an easy way to describe growth and other such biological parameters.

Since age readings are much more difficult to obtain than most other measurements, attempts are usually made to record many attributes for those fish that are age determined. For example, these fish are commonly also weighed, the sex organs are weighed, fat content is recorded as well as other factor which are relevant to the understanding of the stock in question.

The fish population dynamics are fundamentally affected by the age structure of the population, whether or not these ages can be measured. It is therefore essential to consider the basic age effect, and at a later stage one can consider how these effects can be estimated without direct measurements.

3.3.2 Examples

Example 3.2. The figure above shows the hard part of a haddock known as the otolith. The otolith is put into resin and then sliced. The rings are then counted to determine age. Thus, the haddock pictured is a 4 year old.

3.4 Main biological measurements

Typical data on each fish

no	le	wt	sex	mat	age
1	14.5	12.5	0	4	3
2	15.0	13.0	0	5	3
3	14.5	12.0	0	4	3

Must be stored consistently with information on location, gear, time, etc.

3.4.1 Details

When sampling for age, the main biological measurements are usually set up in a simple table, one fish per row. Given the importance of length (le), weight (wt), maturity (mat), sex and age, these data should always be collected for every fish sampled.

It is fairly well known that if age data is available, then length data adds little information to a data set of this form. Further, simple random sampling for ages is in almost all cases better than attempts to stratify and take a fixed number of age readings from each length group, for example. Such a stratification also completely destroys any statistical properties of the data and should be avoided if at all possible (sometimes the fish have already been sorted at the landing site and in this case it may be unavoidable to resort to stratified sampling).

3.4.2 Examples

Example 3.3. The table below contains a subset of the recorded information on 50 capelin which was taken from in a single sample on a research cruise. The entire dataset can be found at:

http://tutor-web.net/fish/fish5101fishsci/lecture30/base.dat

The table contains examples of the main factors which are recorded when fish are taken for age determination. This contains the length of the fish (in cm), its weight (g), sex (0=female, 1=male), maturity stage, and age. A recording of -1 indicates that the corresponding item could not be recorded.

All capelin in this sample were mature, which is seen from the fact that the maturity stage is always greater than 1.

If these data are stored in a rectangular file with a simple header, they can be read into R and the length-weight data plotted using e.g.

```
mat<-read.table("http://tutor-web.net/fish/fish5101fishsci/lecture30
   /base.dat",sep="&",header=T)
plot(mat$Length,mat$Wt,xlab="length(cm)", ylab="weight(g)")
```

3.5 Still other measurements

Other measurements include:

- Liver weight
- Fat content
- RNA/DNA ratios
- Size and number of eggs
- Backcalculated growth
- And others

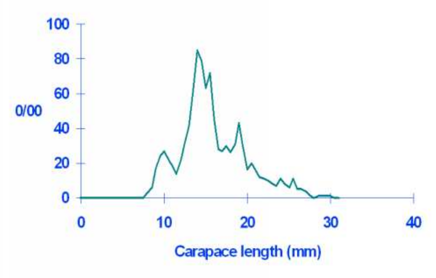
3.5.1 Details

Various other measurements are made and are often informative about the species under consideration. However, few of these measurements have a bearing on the major issues involved in the analysis considered here.

4 Basic analysis of data

4.1 Length distributions

- Basic analysis: Count number in each length cell
- Weight by catch in each stratum - or towing time etc



Carapace length of Northern shrimp *Pandalus borealis* in Icelandic waters.

4.1.1 Details

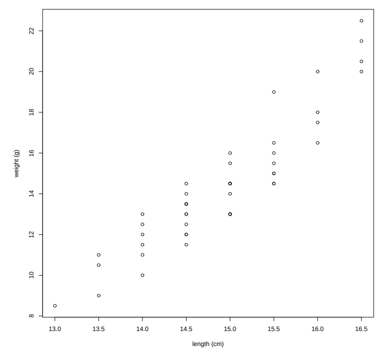
Length measurements are aggregated into length distributions which reflect the number of fish, L_l , in each length group. Usually the length distributions are such that it is possible to identify the youngest, 1-3, age groups but the oldest age groups become indistinguishable.

4.1.2 Examples

Example 4.1. For species such as shrimp, which are difficult to age, there is really no alternative but to measure length for a large number of individuals.

In the case of Northern shrimp, *Pandalus borealis*, in Icelandic waters, it so happens that the species lives for a number of years. Growth is sufficiently fast so that each yearclass has a peak within a length distributions. Thus, yearclasses can be visualized from a length distribution, as seen in the above figure.

4.2 Length-weight relationships



Weight of capelin plotted against length.

4.2.1 Details

There are two principal methods to obtain mean weights. The first is simple, based on taking a substantial number of fish from the catches and computing their mean weight. For example, it is common to weigh a number of boxes and count the contents or to weigh the length samples themselves. Usually, however, attempts are made to weigh individual fish and ideally to weigh the same fish as are later taken for age determination and length measurements. In such cases it is easy to compute the mean weight of fish in the catch. Such samples can also be used to investigate the relationship between weight and length or to compute the mean weight at age directly, particularly if the individual fish are sampled at random.

The more complex method is based on an initial estimate of a length-weight relationship. Such methods must be used when, for technical reasons, it is not possible to weigh individual fish or in some other way to obtain easily the average weight of fish in a sample. The length-weight relationship is found by taking a sample of fish which are individually length-measured and weighed. This usually shows a very good relationship between length and weight.

Definition 4.1. Length-weight relationship estimation:

$$w = \alpha l^{\beta}$$

After such a relationship has been estimated, its stability from year to year is verified. If the relationship is reasonably stable it is possible to use the same relationship for several years, which may result in considerable savings in weighing fish, particularly at sea where considerable sampling is undertaken.

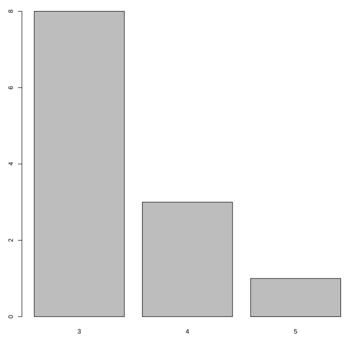
For a given length-weight relationship it is possible to compute the mean weight across all fish sampled in year y

Definition 4.2. Mean fish weight for a given length-weight relationship:

$$\bar{w}_y = \frac{\sum_l \alpha l^{\beta} L_l}{\sum_l L_l}$$

4.3 Age composition

- A stock consists of age groups (cohorts)
- Need to think in terms of cohorts which are caught, grow, mature, spawn, and die.
- If there are age markers (rings) on hard parts, then one can randomly sample the catches to obtain the proportion of fish in each age group.
- Need random samples
- Need good coverage in space and time by gear



4.3.1 Details

The age reading samples can be used to provide a count of the number of fish in each age and length cell. This is usually set up as a length-by-age frequency table where the elements in the table may be denoted with K_{la}.

Definition 4.3. Length-by-age frequency table notation:
K_{la}=table element
a= age of the fish
l= fish length.

For example, the number of fish age 3 and 12 cm long is denoted $K_{12,3}$.

These numbers can be added up within a length group to obtain $K_{l.}$ (The "dot" here, as usual, indicates that summation has been performed over the corresponding variable).

It should be mentioned that if samples taken for age determination are taken at random from the collection of all landed fish then the proportion of fish in each age class can be estimated.

Definition 4.4. Proportion of fish of age a **in year** y:

$$p_a = \frac{K_{.a}}{K_{..}} = \frac{\sum_l K_{la}}{\sum_{la'} K_{la'}}$$

This method of computing proportions is sometimes considered to be undesirable since there may be considerable variation in the otolith samples. In particular it is difficult to obtain adequate samples of the older (and longer) fish unless these are specifically sought out in the sampling, making it non-random. Therefore length distributions are sometimes used to improve the estimate. In this case better estimates may be obtained for the older fish (at the cost of reduced precision for the younger or more abundant age groups), but it is no longer possible to use the age samples alone - they can now only be used in conjunction with the length samples.

4.3.2 Examples

Example 4.2. The above capelin example may be combined into a table which gives the number of fish by age and length, such as in the following table:

http://tutor-web.net/fish/fish5101fishsci/lecture40/age-length-table-from-capelin-otolith.dat

This table shows that the sample contains mostly three year old capelin (the sample was taken in the second part of the season, i.e. during the January-March period). It is also seen that three year old capelin can be from 13 to 16.5 cm long. The example shows that when a year-class dominates a fishery, then the length interval for that year class can reach as far into the upper length groups as an older age group does. The reason is that there are more three year old fish than other year classes. For this reason it is usually desirable to obtain measurements of age as it would otherwise be difficult to see that there is more than one year class in this sample. The last line in the table contains $K_{l.}$ and the rightmost column contains $K_{.a}$.

This table can be obtained in R by using the following commands:

http://tutor-web.net/fish/fish5101fishsci/lecture40/capelin-age-length-from-otolith.r

4.4 Age composition from age and length data

- May have age samples stratified by length
- Need to use length distribution with an age-length key (ALK)

4.4.1 Details

The information from the age samples and length measurements can be used together to compute the mean length and proportions in each age group.

Definition 4.5. Age-length key equation:
In the age samples, length group l comes from the various age groups in such a way that the proportion

$$\frac{K_{la}}{K_{l.}}$$

of length group l is in age group a.

Note 4.1. Tables comprised of the proportion of known age fish in a particular age-length group is referred to as an age-length key (ALK).

It is natural to expect the length distribution to be split into the age groups in the same way. Thus the otolith/scale sample is used to separate the length distribution into age groups and obtain an improved estimate of the number of fish of age a and length l.

Definition 4.6. Age-length distribution equation:
Estimate of the number of fish of age a and length l from length distributions separated by age groups:

$$K'_{la} = \frac{K_{la}}{K_{l.}} L_l$$

Note 4.2. The revised numbers, called the age-length distribution (ALD), are then used to compute the proportion in each age group as well as the mean length at age, as was done earlier.

Note 4.3. An age-sampling scheme which takes, as a basis, some method for taking a fixed (or target) number of fish per length group for ageing is called length-stratified.

Given an average weight it is possible to compute the total number of landed fish by dividing the mean weight into the total landings in tonnes. When the total number landed is known, along with the proportion in each age group, it is possible to compute the total number of landed fish in each age group.

4.5 Catches in numbers at age

To estimate the number of fish caught the following steps are completed:

1. Determine the proportion in each age group from age sampling (p_a)
2. Obtain the total weight landed (Y)
3. Weigh some samples to obtain mean weight of fish in the catches (w)
4. Calculate the total number landed ($C = Y/w$)
5. Use the proportion in each age group to compute numbers caught ($C_a = p_a C$)

4.5.1 Details

When the mean weight has been computed it is also possible to compute the total number of landed fish. This is obtained by dividing the average weight ($\bar{w}_y$) into the total catches in tonnes in year y (denoted Y_y).

Definition 4.7. Total number of landed fish:

$$C_{.y} = \frac{Y_y}{\bar{w}_y}$$

The total count needs to be age determined and this is done by using the earlier obtained numbers on the proportional split into age groups.

Definition 4.8. Age determined total landings

$$C_{ay} = C_{.y} p_{ay}$$

It should be noted that although the year has not been specifically indicated in the age determinations and length distributions, such measurements are always conducted within each year, or even within a part of a year. In fact, attempts are usually made to separate the catches-by-age by gear, seasons, and major oceanic areas. The total catch in numbers by age is then computed by using simple addition.

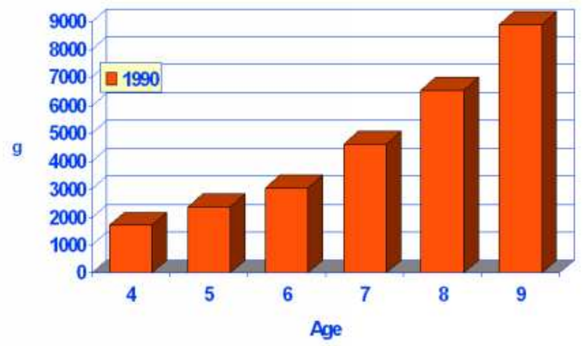
4.5.2 Examples

Example 4.3. These computations are often added to the capelin table, basesl15.dat, resulting in the following expanded table:

http://tutor-web.net/fish/fish5101fishsci/lecture40/expanded-age-length-table-from-capelin-otolith.dat

This example is unusual since the mean length of a five year old (15.5 cm) is lower than the mean length of four year old capelin (16.0 cm). From this it is clear that if mean length at age is to be considered for age groups which are rare in the catches (for example poor year-classes) or are being considered outside traditional fishing grounds, a considerable sampling effort may be needed. In particular, if it is desired to obtain 100 or more five year old capelin to obtain a decent estimate of their mean length, then 50 times more capelin would be required considering the above example, or 2500 capelin.

4.6 Other measurements

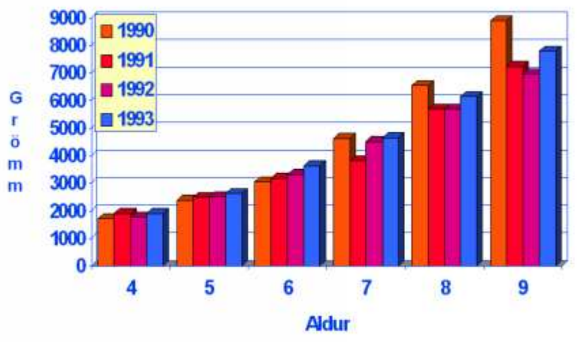


Mean weight at age of cod in Icelandic waters by age groups in 1990. The data was obtained from sampling commercial catches.

4.6.1 Details

Although length measurements and length-age combinations are interesting, it is usually of much greater interest to consider growth in weight and hence weight at age is needed.

4.7 Comparing mean weight at age



I-Cod: Mean weight at age in catch. Repeated measurements permit the comparison of weight at age across years and age groups.

4.7.1 Details

Having obtained weight at age the next step is to obtain a time series of such numbers. They can be compared as a time series for each age group (e.g. for environmentally induced changes in growth), as a growth curve by tracking a cohort, or as histograms of several age groups and years, to obtain information on factors which may be present by year and/or age.

4.8 Mean length at age

Can compute mean length at age from

- Raw data
- Age-length key and age-length distribution

4.8.1 Details

As with the proportion at age, it is simple to compute the average length within each age group.

Note 4.4. The mean length is simply defined as the sum of all lengths divided by the total number of measured fish (within the age group).

If the raw lengths from simple random sampling are available, then the computation of mean length at age is simply computed as the average of the length measurements within each age group.

In the case of data grouped into counts for each age and length cell, things become a bit more complicated. In a given age group, a , there is a total of K_{la} fish measured of length l. Therefore the sum of all their lengths is given with lK_{la}.

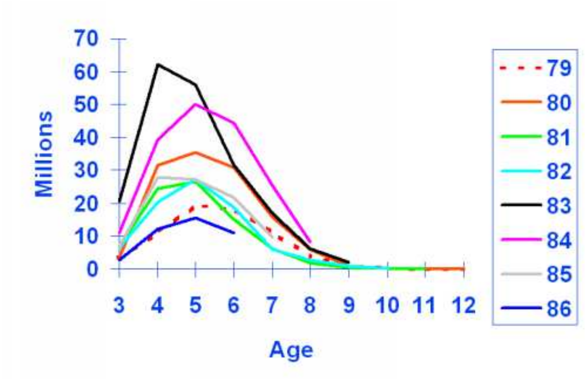
Definition 4.9. Mean length for count data:

$$\bar{l}_a = \frac{\sum_l lK_{la}}{K_{.a}} = \frac{\sum_l lK_{la}}{\sum_l K_{la}}$$

The equation seems more complex than it really is, since this is simply a mean length computed as a sum of lengths divided by the total number. Thus, the numerator is the sum of the contribution by each length group and the denominator is the total count.

When the the data on ages and lengths have been stratified by length group, the age-length table needs to be recast by turning it into an age-length key and corrected with a randomly sampled length distribution.

4.9 Alternate data grouping



Each line in the above figure corresponds to the catches over time from a single cohort.

4.9.1 Details

Although it is useful to consider the catches in number by age in a given year, it is much more informative to track cohorts, i.e. to follow, year by year, the number of fish landed from a given yearclass.

When this is done, and several cohorts are plotted together, a clear picture emerges of the relative abundance of each yearclass, in relation to the abundance of adjacent yearclasses.

These data form the basis of age-based assessment procedures.

4.10 Age composition summary

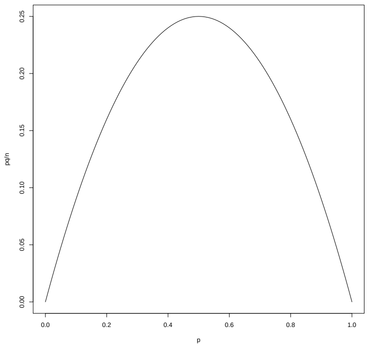
- Age-specific catches are the single most important source of information on mortality in the stock!
- No other measurements can substitute for good age compositions!
- Hence we often try to obtain these through alternate means if there are no annuli
- But if no age readings are available it may be better to use age-based models of length compositions

4.10.1 Details

The usefulness of the age composition of catches can not be overemphasized. These data, when available with some accuracy, provide more information on the state of stock and mortality than any other source.

5 Sampling

5.1 Design issues



The variance vs mean relationship for an estimated proportion

5.1.1 Details

Length distributions are basically count data, as are the number of fish at a give age if simple random sampling is used for age readings. In the case of simple random sampling from an entire population of fish, the number of fish in a given age or length cell should follow a binomial distribution. This implies that the probability of obtaining exactly x fish in a given cell is given by binomial probabilities:

$$p(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

where p is the true population proportion in the cell.Fisheries data rarely satisfy this!

If the binomial distribution applies, the variance of the number of fish in a given cell is npq where $q = 1 - p$ and the mean number(expected value) is np.

Given data, x, the estimated proportion is $\hat{p} = x/n$ and the variance of this estimate is given by pq/n.It should be noted that this variance is never greater than at $p = \frac{1}{2}$.

A common method of obtaining a confidence interval for the proportion is through the use of a Gaussian approximation.

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}\hat{q}}/\sqrt{n}$$

The width of the confidence interval is

$$2z_{1-\alpha/2}\sqrt{\hat{p}\hat{q}}/\sqrt{n}$$

which is never wider than

$$2z_{1-\alpha/2}\sqrt{\frac{1}{2}\cdot\frac{1}{2}}/\sqrt{n}.$$

It follows that if one can guarantee that the latter is no more than δ, then the former will also be smaller.

Commonly, $\alpha = 0.05$ is used and then $z_{1-\alpha/2} = 1.96$.

Some design criteria include:

- The distance between $\hat{p}$ and p should be no greater than δ,
- The confidence interval width should be no more than δ.
- The standard error of $\hat{p}$ should be no more than δ.

Confidence interval width criterion:

$$\begin{aligned} & 2z_{1-\alpha/2}\sqrt{\hat{p}\hat{q}}/\sqrt{n} \leq \delta \\ \Leftrightarrow \quad & n \geq (2z_{1-\alpha/2}\sqrt{\hat{p}\hat{q}}/\delta)^2 \end{aligned}$$

Now note that

$$n \geq \left(2z_{1-\alpha/2}\sqrt{\frac{1}{2}\cdot\frac{1}{2}}/\delta\right)^2 \Rightarrow n \geq \left(2z_{1-\alpha/2}\sqrt{\hat{p}\hat{q}}/\delta\right)^2 \Rightarrow 2z_{1-\alpha/2}\sqrt{\hat{p}\hat{q}}/\sqrt{n} \leq \delta$$

and it follows that the confidence interval is sufficiently tight if n is chosen to satisfy $n \geq \left(z_{1-\alpha/2}/\delta\right)^2$.

Distance between $\hat{p}$ and p criterion:

$$n \geq (z_{1-\alpha/2}\frac{1}{2}/\delta)^2$$

Standard error of $\hat{p}$ criterion:

$$n \geq (\sqrt{\frac{1}{2}\cdot\frac{1}{2}}/\delta)^2 \Rightarrow n \geq (\sqrt{\hat{p}\hat{q}}/\delta)^2 \Rightarrow \sqrt{\hat{p}\hat{q}}/\sqrt{n} \leq \delta$$

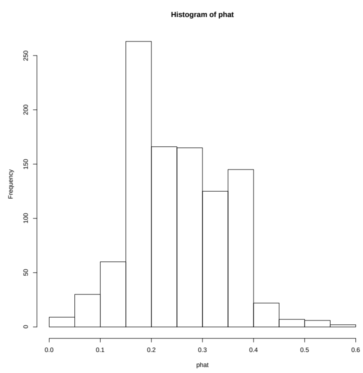
5.2 Simulating sampling schemes

Want to estimate a proportion p

Sample independently individual fish (in group=1; or not=0): get binomial distribution for x=number of positives

Estimated proportion: $\hat{p} = x/n$

Known expected value and variance: $E[\hat{p}] = p$, $V[\hat{p}] = np(1-p)$



5.2.1 Details

The simplest assumptions are that one is estimating a single proportion, p, by counting 0/1-values in n independent experiments. In this case the number of positive outcomes, x, is an observation from a binomial distribution with a probability p of a positive outcome in n trials. The unknown proportion p is estimated with $\hat{p} = \frac{x}{n}$.

If we sample independently individual fish, this is is how the proportion in a given length group is obtained. Similarly, within a catch we could estimate the proportion of a given species by taking a random sample of n fish within the catch and the same would apply.

The statistics of this are very well known with the long-term average value of $\hat{p}$ being p and the true variance is $p(1-p)/n$, but only if the assumptions of independence hold true.

5.2.2 Examples

Example 5.1. In R it is very easy to simulate the effects of different sampling schemes and there are several ways to do this.

Consider a red/green marble experiment, where we first define the sample size (number of marbles to be drawn) and set up the sequences of red and green marbles along with the entire set:

http://tutor-web.net/fish/fish5101fishsci/lecture50/sampling-scheme-simulation-1.r

Example 5.2. Consider an evaluation of whether the confidence interval for a proportion has the correct coverage probability.

http://tutor-web.net/fish/fish5101fishsci/lecture50/ci-coverage-probability.r
A typical session would look as follows:

http://tutor-web.net/fish/fish5101fishsci/lecture50/simulating-sampling-schemes.pdf

5.3 Correlation issues

When correlation issues occur the model is usually wrong

Primary issue: Intra-haul correlation, cf Pennington and Volstad

Fish within a station s are more similar than across stations.

A model of the process needs to include a (random) station effect: $y_{sj} = \mu + \alpha_s + \varepsilon_{sj}$

Here, y_{sj} could be the length of fish j at station s.

The resulting correlation between fish at the same station is the **intra-haul** correlation

6 Overview

6.1 Summary and exercises

Important fisheries issues

- Must monitor the resources - need good sampling schemes
- Need to look at the overall picture, not just current stock size
- Must have a basic biological knowledge
- Assessments are harder when less data is available

6.1.1 Assignment

Assignment 6.1. Fish population dynamics

Answer at least 8 quiz questions correctly from each lecture in this tutorial.
Download and subsequently read into R data on length, maturity, weight, age etc. This can be done with a command of the form

```
fish<-read.table("http://www.hi.is/~gunnar/kennsla/alsm/data/set111.
dat",header=T)
```

Look at the data and see where there are missing values. Delete rows where ages or lengths are missing. Check to see whether there are other obvious problems.

Tasks in assignment 1:

- Set up the length distribution, first in counts and then plot it as percentages per length group.
- Compute the mean length at age from the raw data. Plot the mean length against age.
- Fit a model (von Bertalanffy) (a) to the mean length at age data. (b) to the raw data from individual fish.
- Plot the weight against length. Suggest a transformation to linearity. Check whether there are data problems and reduce the data set if needed.
- Fit the length-weight relationship $w = \alpha l^{\beta}$ using an appropriate transformation. Transform back to obtain α and β.
- Set up the counts in the age-length table. Compute the age-length key.

```
Columns in data set

name   Icelandic                           English
----   ----------------                    ---------------
safn   Númer gagnasafns                    Data set number
teg    Fisktegund                          Species
ar     Ár                                  Year
man    Mánuður                             Month
reit   Reitur                              Statistical rectangle
dag    Mánaðardagur                        Day within month
dypi   Dýpi (m)                            Depth
vf     Veiðarfæri                          Fishing gear
le     Lengd fisks (cm)                    Length of fish
ky     Kyn (1=hængur)                      Sex
kt     Kynþroski (1=ókynþroska)            Maturity stage
aldur  Aldur, ár                           Age in years
osl    Óslægð (lifandi) þyngd (g)          Ungutted (live) weight
sl     Slægð þyngd (g)                     Gutted weight
li     Lifrarþyngd (g)                     Liver weight
```

Note that one can rename the columns with e.g.

```
names(fish)<-c("set","species","yr","mon","square","day","depth","
   gear","le","sex","mat","age","whole","gutted","liver")
```

Further practicals

Tasks

In this practical you will learn how to do the following in R:

Length distributions
Age–length tables
Age–length keys
Age–length distributions
Proportion at age
Mean length at age
Weight–length relationship
Catch in numbers at age
Summary plots

Length and age data

With length and age data it is important to be able to know how to create length distributions, age–length distributions etc It is also important that you practice looking at the data – don't just type in the commands, look at what they do. Also, don't just do the plots I ask for, try other plots as well.

The most important R commands used here aggregate data and use the dimension names of objects. There are examples of all of these in the Introduction to R.

For some commands an example has been given with a small dataset to try first. When using a new command, or trying to do something complex it can be helpful to first use a small dataset where you know what the answer should be or where it is easy to see exactly what has been done.

Some of this may look quite complicated – it just takes longer to explain than do. It's important, however, that you use a text editor to keep a record of your commands so you can easily replicate the process later.

The data are available at:
http://tutor-web.net/fish/fish5101fishsci/lecture60/age.dat
http://tutor-web.net/fish/fish5101fishsci/lecture60/length.dat

- Copy the data into a text editor – remember it is easier if you run R in the same directory as the file you are reading in.
- Read the data into R using read.table (eg `adata <- read.table("age", header=T)`) which creates a data.frame. If R is not running in the same directory you will need to provide the path of the file.

The columns are:

station	a unique number for each haul/tow
rectangle	position: each rectangle is one degree longitude and half a degree latitude
species	code 1 is cod
le	length in cm
age	age in years
wgt	weight in grams
number	the number at that length OR the number at that age and length

The age data are a subset of the length data, so all lengths in the age table should also be in the length table.

Length distributions – from a vector of length data
With the age data we have only one fish for each row. With data in a format like this a length distribution can be created using `table`.

- Example: Given a vector of length data: `x = sample(5:12, 20, replace=T)`

  ```
  sort(x)
  table(x)
  hist(x)
  ```

- Create a vector containing the length data from the age file. (eg `le <- age$le`)
- Plot it and tabulate it (as for the example). `table` calculates the length distribution.

As these data are sparse – only zero or one fish for most lengths – it is maybe better to aggregate.

- Group into 5cm length classes. `le5 <- floor(le/5)*5+2.5` What does the `2.5` at the end do?
- How would you do this for 10cm groups?
- Plot histograms to look at the data.
- To create a length distribution you need to count the number of fish in each length group. Create the length distribution.

As the length distribution should be in even groups we have some missing. With a large dataset this is less likely to be a problem but there still might be gaps at the ends of the distribution.
The gaps in a vector can be filled in as follows:

- Create a vector the length the length distribution should be and name the cells of the vector. eg

  ```
  rnames <- seq(min(le5), max(le5), 5)
  ldist <- rep(0, length(rnames))
  names(ldist) <- paste(rnames)
  ```

- Then, calculate the length distribution from the data and put it into the empty vector. This substitutes the elements of `tmp` into `ldist` where `names(tmp) = names(ldist)`.

  ```
  tmp <- table(le5)
  id <- names(tmp)
  ldist[id] <- tmp
  ```

- This can be done in fewer lines by combining commands:

  ```
  rnames <- seq(min(le5), max(le5), 5)
  ldist <- rep(0, length(rnames))
  names(ldist) <- paste(rnames)
  tmp <- table(le5)
  ldist[names(tmp)] <- tmp
  ```

- Plot the length distribution.

  Line plot: `plot(as.numeric(names(ldist)), ldist, type="l")`

  Barplot: `barplot(ldist)`

- Calculate and plot the length distribution as proportions – the number in each length class divided by the total number. (Use `sum` to calculate the total.)

- Practie calculating length distributions with different levels of aggregation.

Length distributions – from a matrix of length and count data
With the data in the length file we can have more than one fish per row which means we cannot use `table` and use `tapply` instead.

- Example: Create a small dataset temp and use `tapply`. Then plot the length distribution.

```
le <- c(seq(20,40, 5), 25, 35)
num <- c(4,7,9,6,2, 5,3)
temp <- data.frame(le = le, num = num)
temp.tab <- tapply(temp$num, temp$le, sum)
```

- To create a length distribution from the ∼lorna/length data you need to sum the number of fish in each length group. (eg `l.dist <- tapply(ldata$number, ldata$le, sum)`)

- Plot the data (`barplot`). Do these data need to be aggregated into wider length groups?

- If you want to aggregate the length data you can store them as another column in the data.frame. eg `ldata$le5 <- floor(ldata$le/5)*5+2.5`

- Do you have the same problem with empty length cells? If so, fix the length distribution.

- Plot the length distribution as before.

- Calculate and plot the length distribution as proportions.

Age distributions

- Create a barplot of the age distribution using the methods you have learned for length distributions.

- Calculate the proportion at age.

Age–length tables
Using the age data we can tabulate using 2 columns to create an age–length table (ALT) a 2 dimensional table.

- Tabulate the data (eg `alt <- table(age$le2, age$age)`)

- Try with with different length aggregations.

- If you have gaps in the length cells how can you create an even table?

- This is done in a similar way as for the length distribution – but with a matrix. As there are 2 dimensions `dimnames` is used instead of `names`.

  First create a matrix of zeros with dimnames of all possible values of age and length.

  ```
  age.vec <- min(age$age):max(age$age)
  le.vec <- seq(min(age$le5),max(age$le5),5)
  alt <- matrix(rep(0, length(age.vec)*length(le.vec)),
  ncol=length(age.vec))
  dimnames(alt) <- list(le.vec, age.vec)
  ```

  Then calculate the age length table and substitute the age length table into the empty matrix.

  ```
  alt.tmp <- table(age$le5, age$age)
  alt[dimnames(alt.tmp)[[1]], dimnames(alt.tmp)[[2]]] <- alt.tmp
  ```

Example: This small example may be clearer. Here, `tmp2` is a larger matrix of zeros and the values of `tmp` are substituted into it.

```
tmp <- matrix(c(2,4,2,4,2,4), byrow=T, ncol=2)
dimnames(tmp) <- list(c(2,4,6), 1:2)

tmp2 <- matrix(rep(0,10),byrow=F, ncol=2)
dimnames(tmp2) <- list(2:6, 1:2)

tmp2[dimnames(tmp)[[1]],dimnames(tmp)[[2]]] <- tmp
```

Operations on the rows and columns of a table
The length and age distributions of the age data can be calculated directly from the age–length table.
To sum the columns (ages)
`apply(alt, 2, sum)`
To sum the rows (lengths)
`apply(alt, 1, sum)`

- Compare these with the number at length and at age you calculated from the original data.

Functions other than sum can be used in `apply` eg mean, median.

Age–length distributions and mean length at age
Mean length at age can be calculated directly from the age data using `tapply`.

- Calculate mean length at age from the age data.

- Calculate the corresponding variance.

In length stratified sampling schemes age samples are taken from a fixed number of fish by length group. This can cause problems in calculating mean length at age directly from the age data. Even without such a stratification scheme using the length data (which is more abundant) along with the age data provides a better estimate of mean length at age.

The ALT is used to assign an age to fish with only length data. To do this, the ALT is transformed into an age length key (ALK) i.e. each row represents the proportion of fish at each age for a length group eg the age distribution of fish of between 50 and 54 cm. To do this each row of the ALT is divided by the total number of fish in that row ie each column of the ALT is divided by the total length distribution of the table.

The most reliable way to do this is to create a matrix of the same dimensions as the ALT.

- Calculate the length distribution of the ALT
  eg `ld.tmp <- apply(alt, 1, sum)`

- Create a matrix where each column is the length distribution.
  eg `ld.mat <- matrix(rep(ld.tmp, ncol(alt)), byrow=F, ncol=ncol(alt))`

- Create the ALK by ALT/LD – `round` will help if you want to look at it to fewer decimal places.
  eg `alk <- alt/ld.tmp`

- As the ALT contains many zeros, these are now missing values (`NA`). In the ALK these should be changed back to zero. To substitute `0` for every `NA`:
  eg `alk[is.na(alk)] <- 0`.

The ALK is then multiplied by the length distribution from the length samples. Again you need to create a matrix from the length samples. The lengths of the length distribution must match those of the ALK. To do this the `names` of the length distribution must match the length `dimnames` of the age-length key, eg `ld <- ld[dimnames(alk)[[1]]]`.

Age distribution from the ALD

- Replicate the length distribution as before to create a matrix with the correct number of columns (and rows) called here LD – the lengths of LD must match those of the ALK.

- Calculate the ALD by ALK*LD.

- Calculate the number at age from the ALD. (`apply`)

- Calculate the proportion at age.

- How does the proportion at age from the ALD compare with that from the age data alone?

Mean length at age from the ALD
To calculate mean length at age from the ALD. The ALD is a matrix of the number of fish in each age–length cell.

- Multiply the ALD by the length of each cell: the lengths are `as.numeric(dimnames(ald)[[1]])` by:

  replicate the lengths to create a matrix.

  multiply ALD by the length matrix to create a new matrix ALD2

- Sum the columns of ALD2 to calculate the total length of all fish by age. Divide this by the total number at age.

How do the calculations of mean length at age directly from the age data and from the ALD compare?

Sourcing a file
Write a file to do all the necessary steps and run the file in R using eg `source("calc.r")`.

The steps are:

Read in length data
Read in age data
Calculate the length distribution from the length data.
Calculate the age-length table from the age data.
Calculate the age-length key.
Calculate the age-length distribution.
Calculate mean length at age.

To write to the screen from a sourced file it is necessary to use `print(`*objectname*`)` . Normally in R you have just typed *objectname*.

Changing the range of age and length data
In the method described above, the length distribution was reduced to the dimensions of the ALK. An alternative method is to calculate the length distribution only from lengths in the same range as the age data. As the age data are a subset of the length data there will often be smaller or large fish in the length than age data.

The length data can me modified by deleting values outside the range of the age data or by changing the lengths to the smallest/largest in the age sample. The command `ifelse` can be used in either case.

Some examples of `ifelse` are:

```
x <- 1:10
ifelse(x<5, 0, x)
```

if x $<$ 5 then x = 0, else x = x `ifelse(x==5, 0, x)`

To use `ifelse` to delete certain values the values outside the length range are changed to `NA` which means missing and then deleted, eg:

```
mina <- min(adata$le)
maxa <- min(adata$le)
ldata$le <- ifelse(ldata$le<mina, NA, ldata$le)
ldata$le <- ifelse(ldata$le>maxa, NA, ldata$le)
ldata <- ldata[is.na(ldata$le),]
```

Alternatively, `ifelse` can be used to change the values to the minimum or maximum in the age sample, eg

```
mina <- min(adata$le)
maxa <- min(adata$le)
ldata$le <- ifelse(ldata$le<mina, mina, ldata$le)
ldata$le <- ifelse(ldata$le>maxa, maxa, ldata$le)
```

Length–weight relationship
The age data in www.hafro.is/∼lorna/age contains length and weight data.
The length–weight relationship is typically written as:

$$W = a \times L^b$$

which can be rewritten as:

$$\ln W = \ln a + b \times \ln L$$

- Plot weight against length.
- Do any of these data look wrong? Outlying points can be removed as follows:

```
plot(data$x, data$y)
identify(data$x,data$y)
```

  click on the 'bad' point with the left mouse button - it will be numbered

  if you want to identify more points continue to click with the left and when you want to finish click with the right.

  the number of all the identified plots will then be in your R window. You could also have done:

  `id <- identify(data$x,data$y)`

- To remove the rows of data containing the outlying data points and create a new dataset:

  `data2 <- data[-id,]`

  If you didn't put the points into an object you can still do this using the information printed into the R window, e.g `data2 <- data[-c(3,6),]`

- Plot the log transformed data (natural logarithm). Do you have any outliers? Relabel the x and y axes with appropriate text.
- Fit a linear regression model to the log transformed data. (eg `fit <- lm(log(y) ~ log(x))`
- Using `fit$coef` what are the values of *a* and *b* in the first equation above?
- Use `summary(fit)` to look more at the fitted model.
- Plot model diagnostic plots using:

```
par(mfrow=c(2,2))
plot(fit)
```

Catch at age
Catch at age can be calculated using a combination of length data, the weight–length relationship, proportion at age and total catch.

- Using the length data from the length file (the age data are only a subset of it) and the weight–length relationship, what is the mean weight of fish in the dataset? eg

  ```
  mean(a*le^b)
  ```

  The fish were weighed in grammes so convert this into kg.

- If the total catch is 1000 kg how many fish were caught?

- Using the calculated proportion at age, what is the catch in numbers at age? (eg `agep*num`)

Plotting
Use boxplots to summarise the datasets. eg
Compare length distribution by rectangle.
Compare length distribution by age.
Compare number of fish by rectangle.

```
boxplot(le ~ rectangle, data = ldata)
boxplot(le ~ age, data = adata) boxplot(num ~ rectangle, data =
ldata)
```

Try some other plots.