

ЭЛЕМЕНТЫ ДИАГНОСТИКИ

Диагностика чего ?

- Модели ?
- Свойств данных ?
- Результата ?

Примем, что свойства модели уже известны: она протестирована на модельных данных, доказано, что она способна делать то, что продекларировано её авторами, по крайней мере, в условиях умеренной зашумленности данных и небольших отклонений от основных гипотез

Иногда «первичное» тестирование моделей также называют диагностикой.

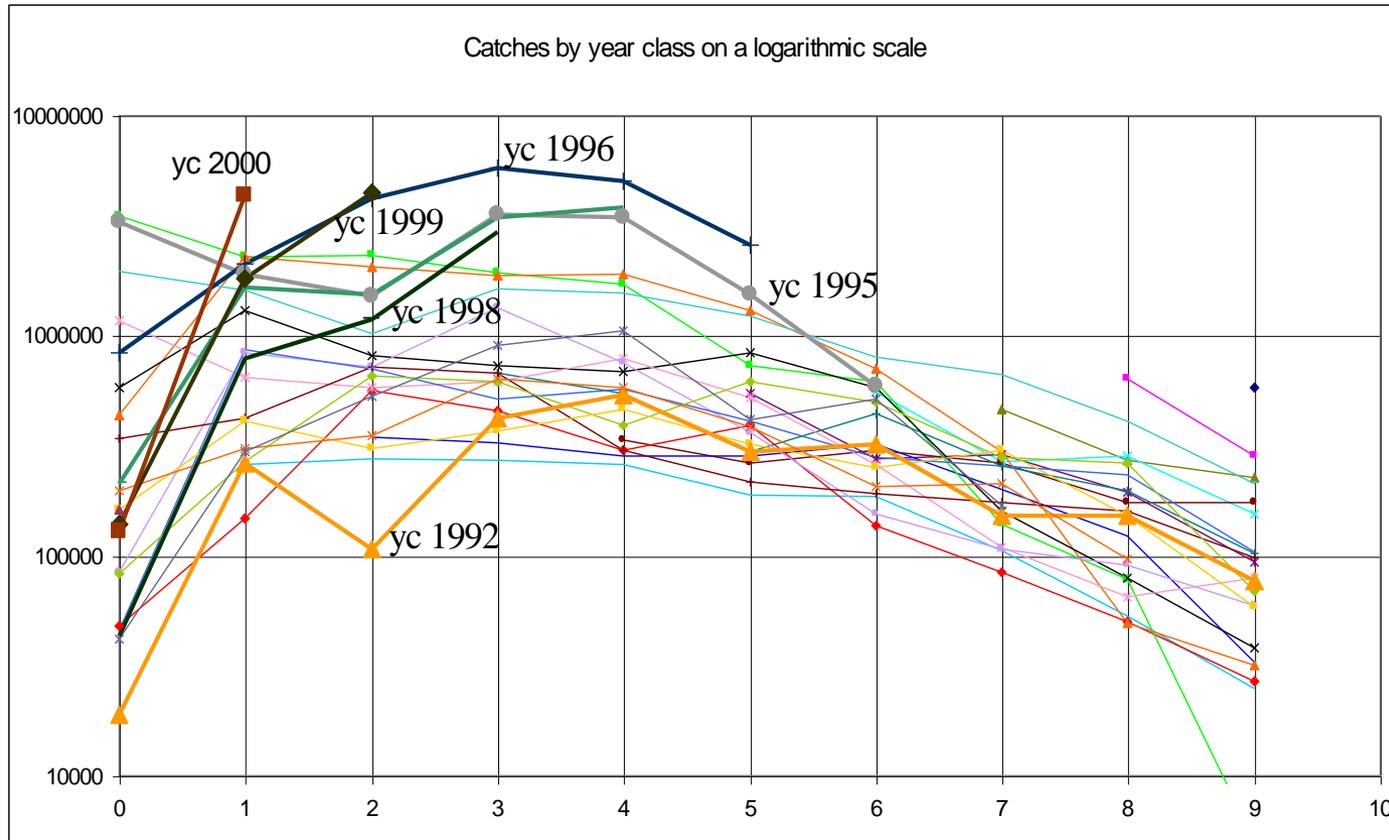
Мы будем понимать под диагностикой выяснение свойств решения как отражение особенностей взаимодействия конкретных данных с конкретной моделью с целью:

- выявления проблемных мест
- определения путей их устранения

С этой задачей творчески имеют дело все специалисты в области оценки запасов, поэтому замечания и дополнения весьма приветствуются

*Первичный анализ данных также важен (чтоб
заранее представлять с чем имеем дело)*

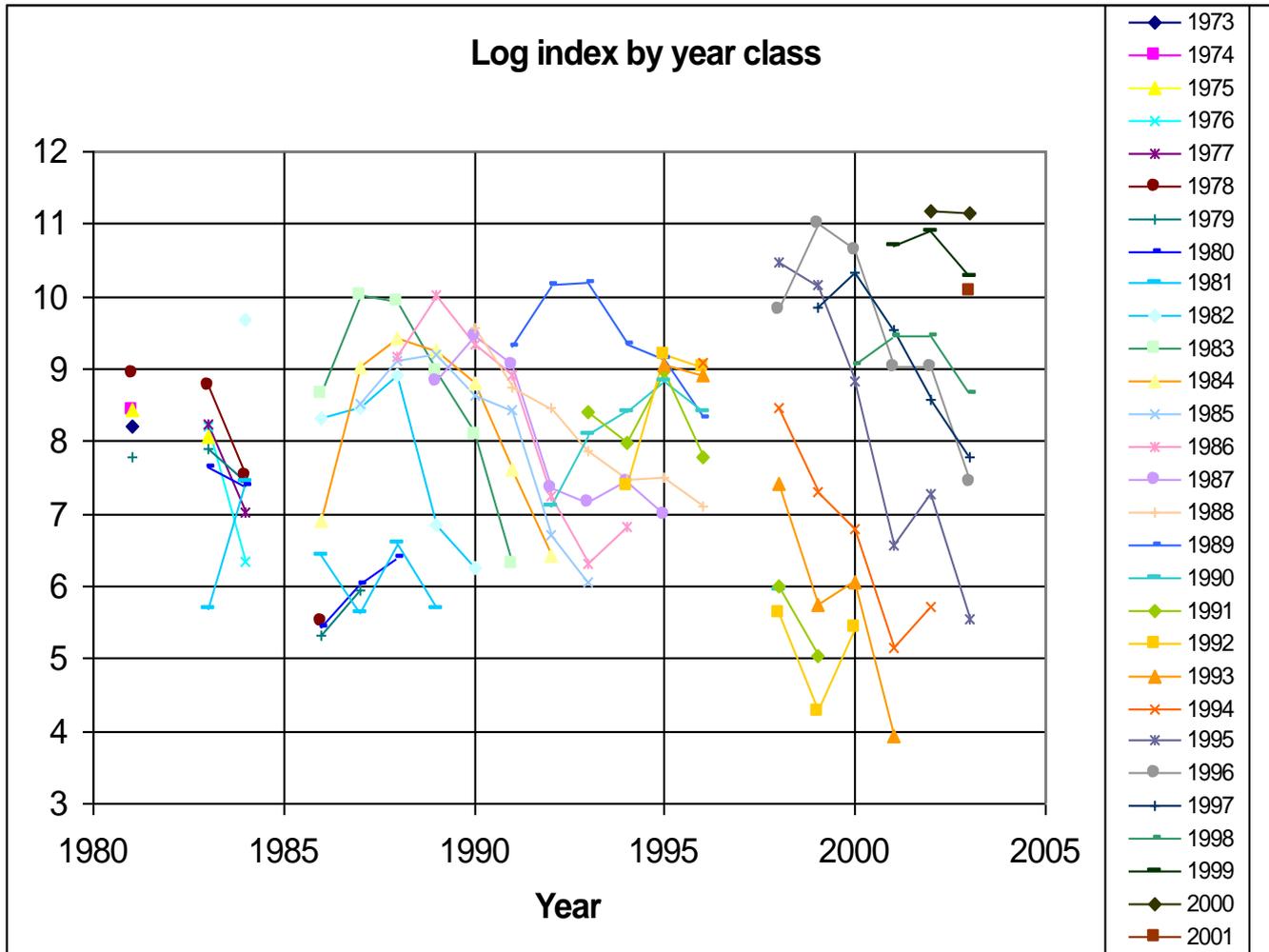
анализ логарифмов уловов по поколениям



-Можно посмотреть на устойчивость селективности

-видно, например, что поколение 1992 г.р. имеет сильную аномалию в уловах в младших возрастах и может быть разумным иметь более высокие остатки для этого поколения. Это означает, что общая оценка в меньшей степени основана на этом поколении

Аналогично – для съёмок

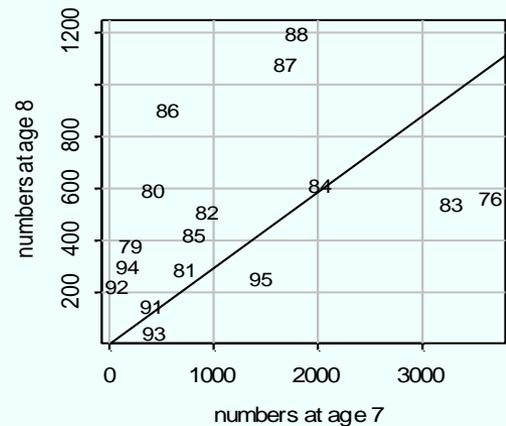
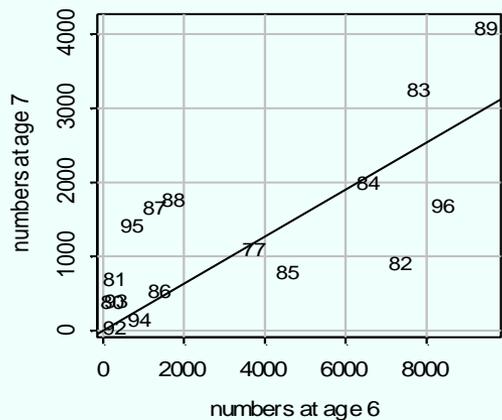
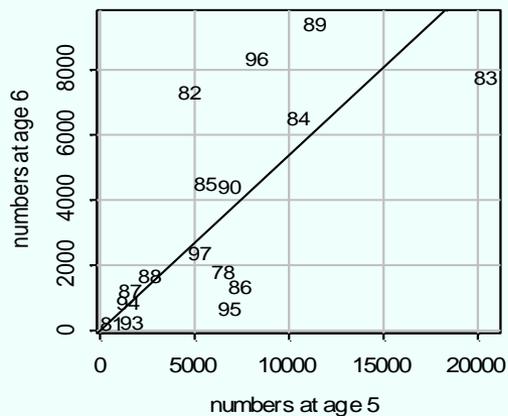
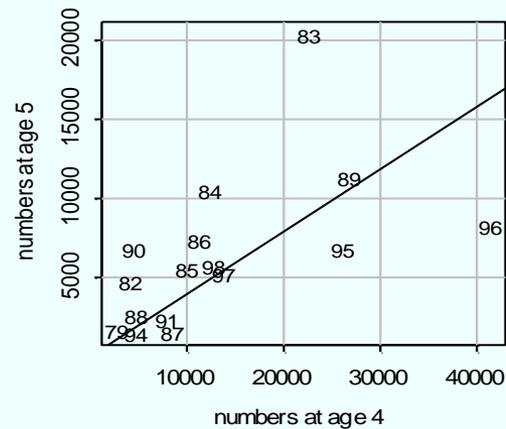
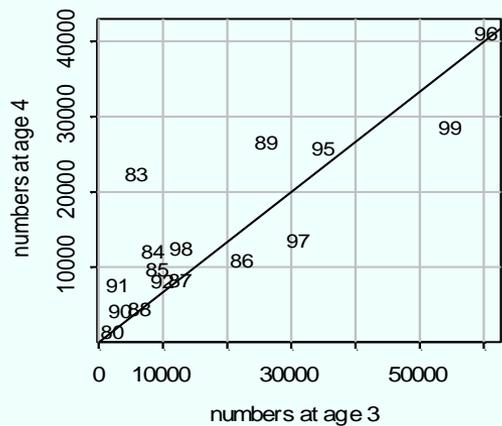
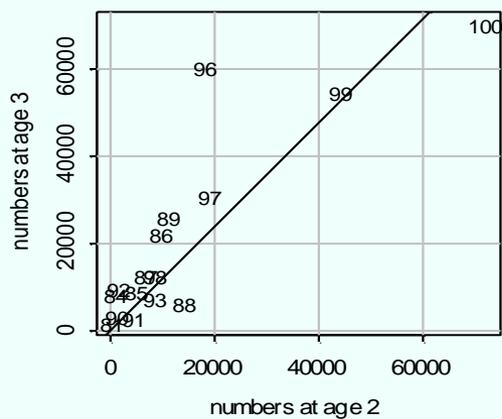


Можно увидеть
Сильные поколения

Колоколообразная
Форма говорит о
низкой улавлива-
емости в младших
возрастах

Has the mortality increased?
Year effects

Анализ корреляции внутри съёмки: $U(a,y)$ vs. $U(a+1,y+1)$

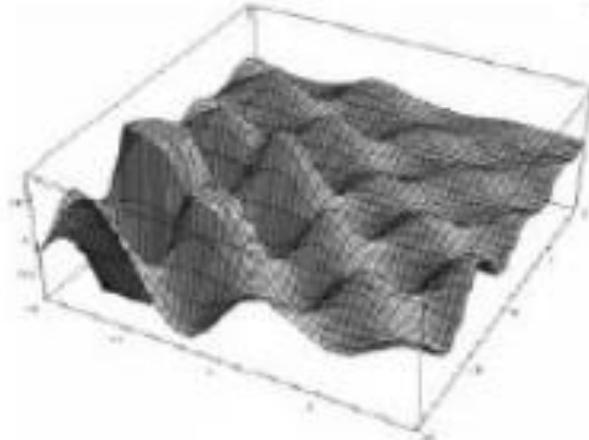


Итак, посмотрели на данные:

- *увидели, что они, как всегда, не идеальны, однако других нет;*
- *подумали о выборе модели и её опций;*
- *применили её и получили некий результат.*

И всё ?

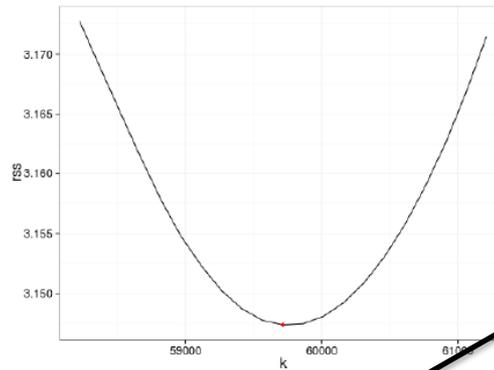
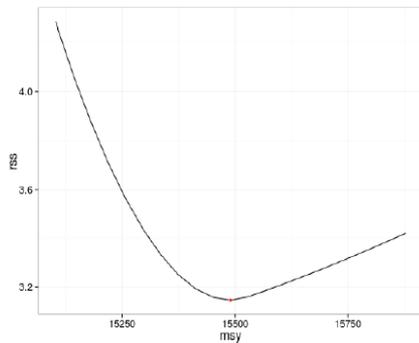
Нет. Всё только начинается.



Прежде всего, надо убедиться, что мы взяли «решение» в минимуме провозглашенной целевой функции модели, т.е.:

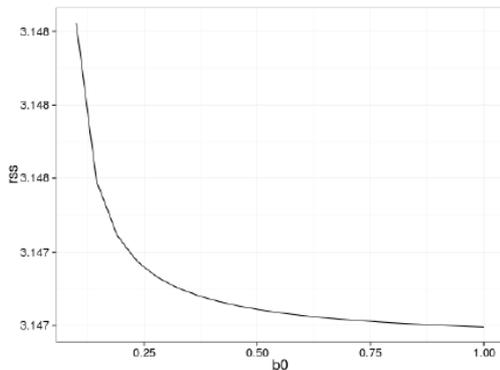
- «доехала» ли минимизационная процедура до минимума, хотя бы – локального?*
- Есть ли невырожденный минимум вообще?*
- Единственен ли он?*
- Если локальных минимумов несколько, то который из них глобальный? В него ли мы пришли?*
- Разумные ли оценки параметров имеем в глобальном минимуме или лучше выбрать, пусть локальный, но дающий оценки, имеющие биологический смысл? (актуально для сильно зашумленных данных с аутлаерами)*
- Отчетлива ли локализация минимумов и т.д.*

Если количество параметров модели обозримо, то:
- «просканировать» поверхность целевой функции, меняя параметры с достаточно мелким шагом, и представить её проекции относительно параметров



Производственная модель с тремя параметрами:

Минимума нет, т.е. нет и решения.

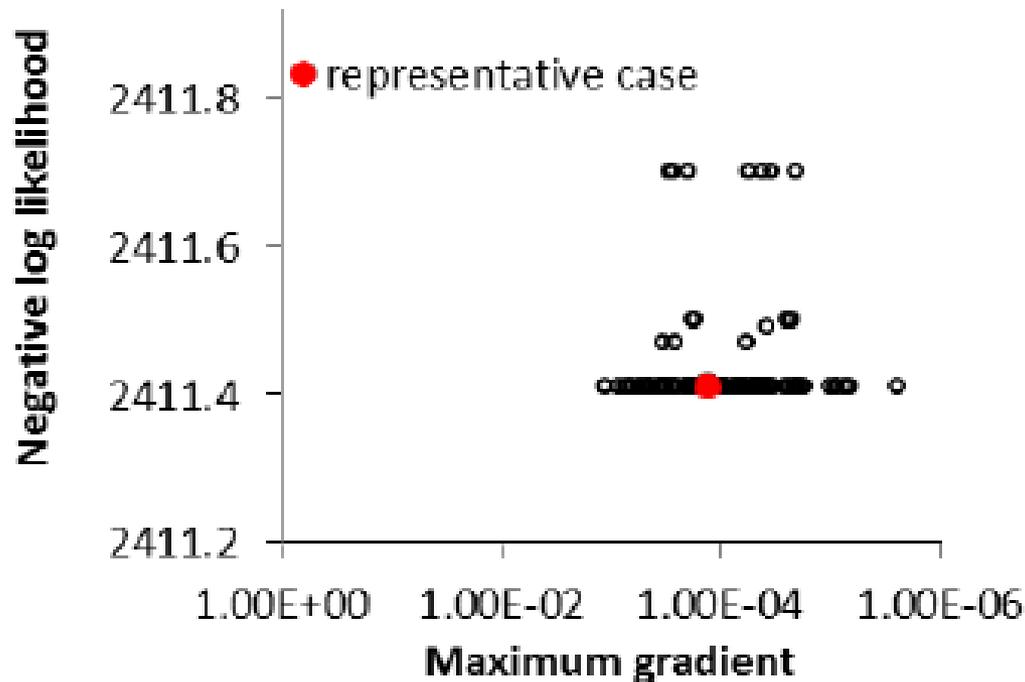


Что делать?

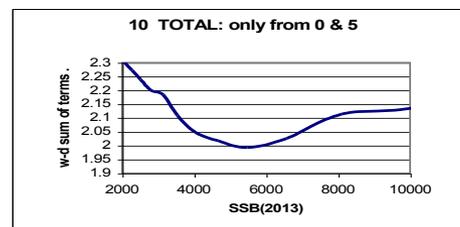
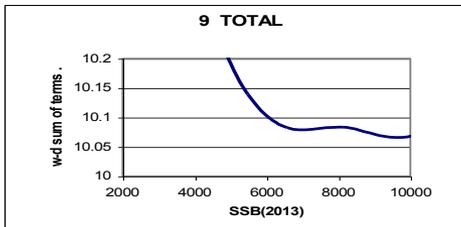
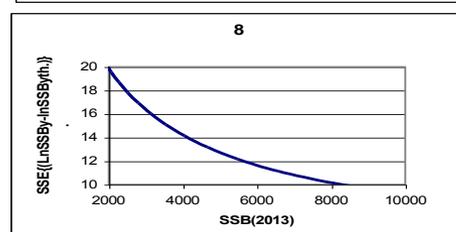
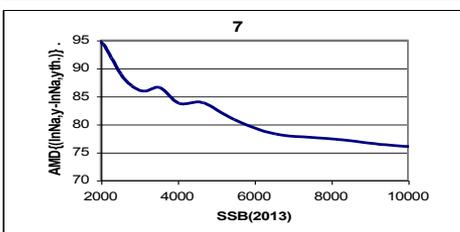
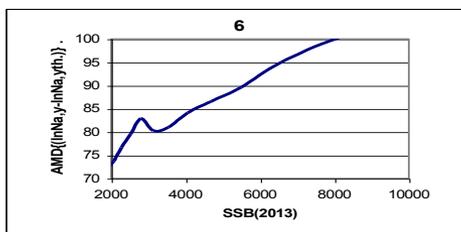
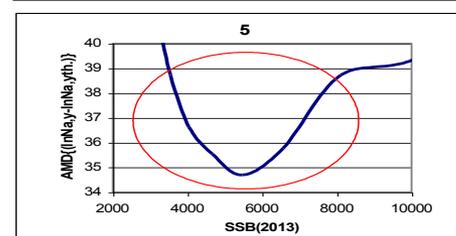
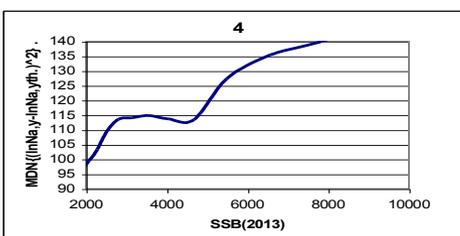
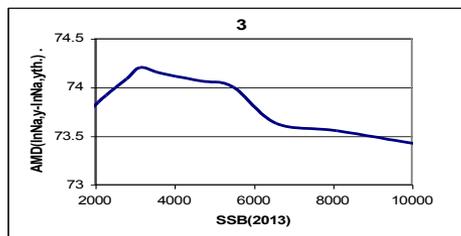
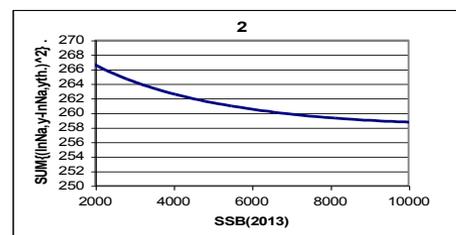
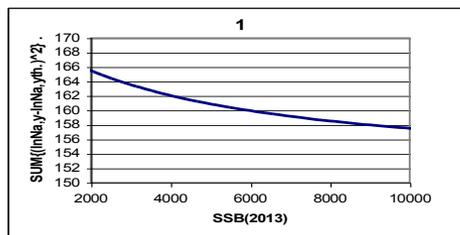
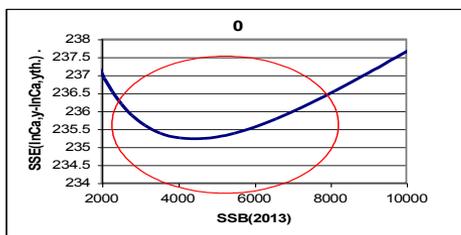
- Переформулировать модель
- Изменить вид целевой функции
- Найти и откорректировать аутлаеры в данных
- Или всё вместе, до победы.

Если параметров очень много и (или) программа не позволяет провести сканирование, то хотя бы

- Провести многократное оценивание, стартуя из разных начальных приближений*
- Варьировать параметры минимизационной процедуры и сравнить конечные значения целевой функции*



Если целевая ф-я модели многокомпонентна, то следует провести анализ информативности данных (профили компонент Ц.Ф.)



Данные

- возрастной состав мировых уловов
- съемки на нерестилищах у побережья Норвегии (съемка 1);
- съемки в зимовальных районах в районе Вест-фьёрда в ноябре-декабре (съемка 2);
- съемки в зимовальных районах в районе Вест-фьёрда в январе (съемка 3);
- майская съемка молоди сельди в Баренцевом море (съемка 4);
- съемка в нагульных районах Норвежского моря в мае (съемка 5);
- совместные российско-норвежские экосистемные съемки (съемка 6);
- съемки - 0-группы (съемка 7);
- данные по личинкам и как индекс биомассы нерестового запаса - «съемка 8»;

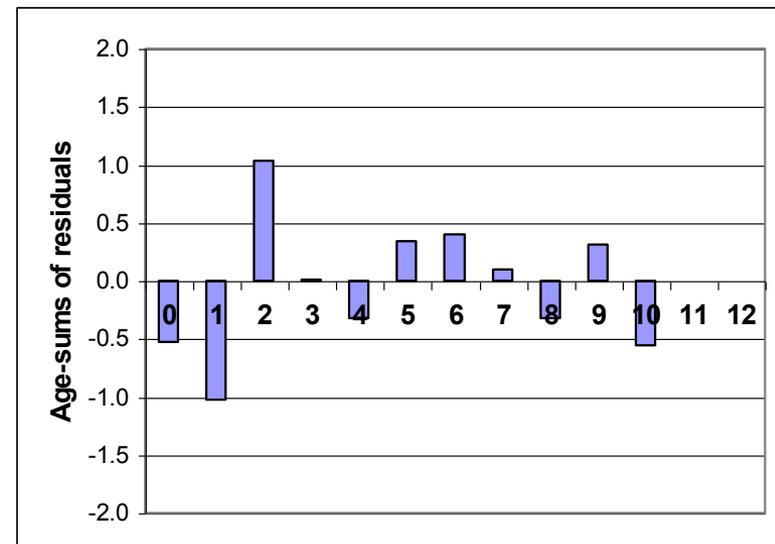
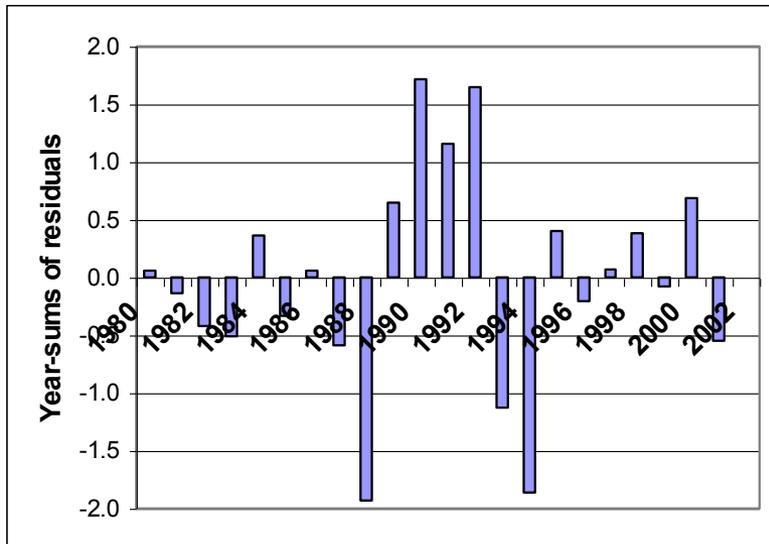
Если сигналы от разных видов данных (положения минимумов соотв. компонент Ц.Ф.) противоречивы, а от некоторых сигнала нет вовсе (нет минимума), то:

- Попробовать улучшить (сблизить) сигналы путем выбора других (более робастных) целевых функций и (или) опций модели*
- Те виды данных, которые так и не дадут вразумительный сигнал, лучше не использовать вовсе*
- Подумать о весовых коэффициентах при компонентах Ц.Ф.*
- Вернуться к анализу входных данных: выявить и уничтожить (или сгладить) аутлаеры с помощью, например, робастной винзоризации*
- Начать всё с начала.*

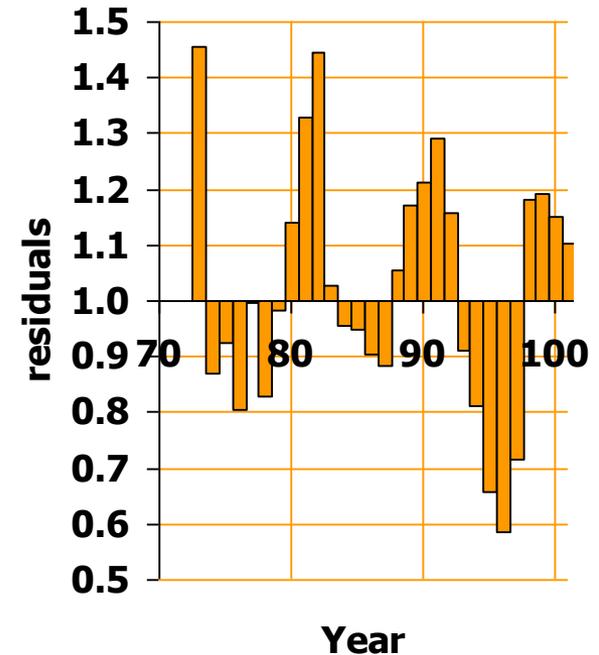
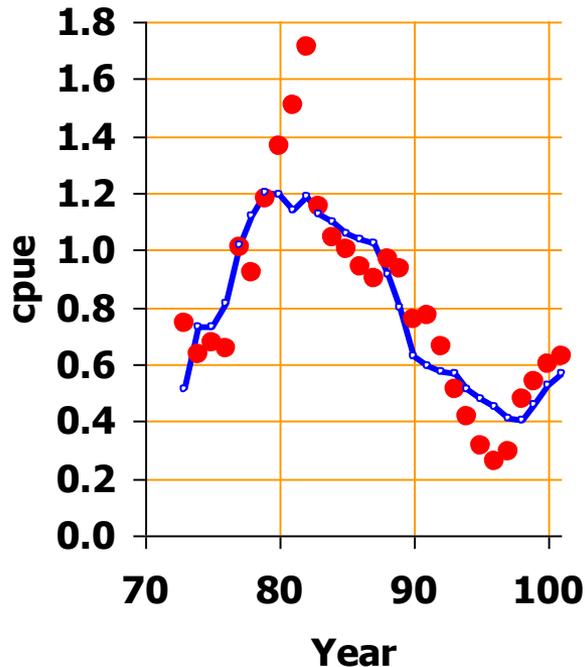
Кое о чём может сказать анализ остатков

1. Суммы остатков (по годам и возрастам)

- «смещенность» описания*
- автокорреляции*



*Или только по годам, если возрастной структуры нет
(как в производственных моделях)*

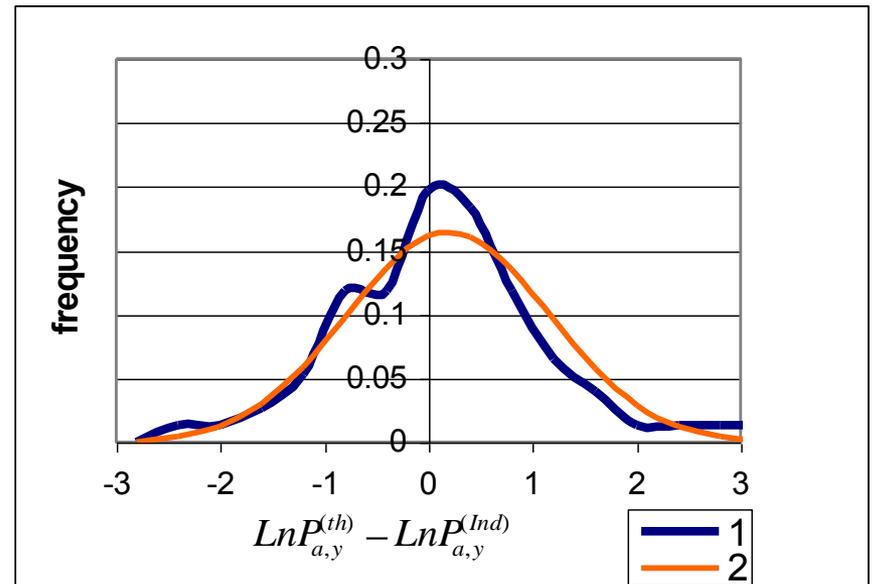
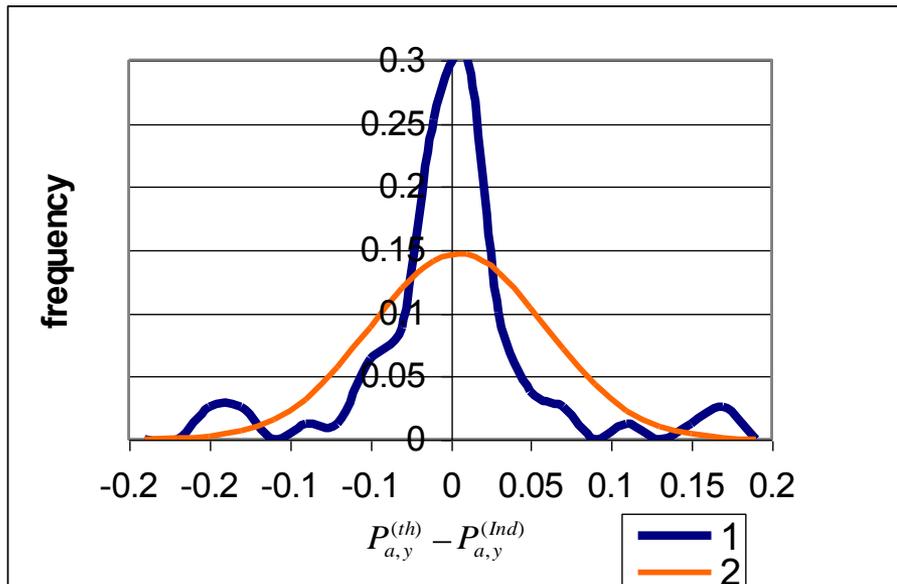


*Некоторые процедуры оценки параметров позволяют
искать только «несмещенные» решения.*

2. Анализ характера распределения остатков

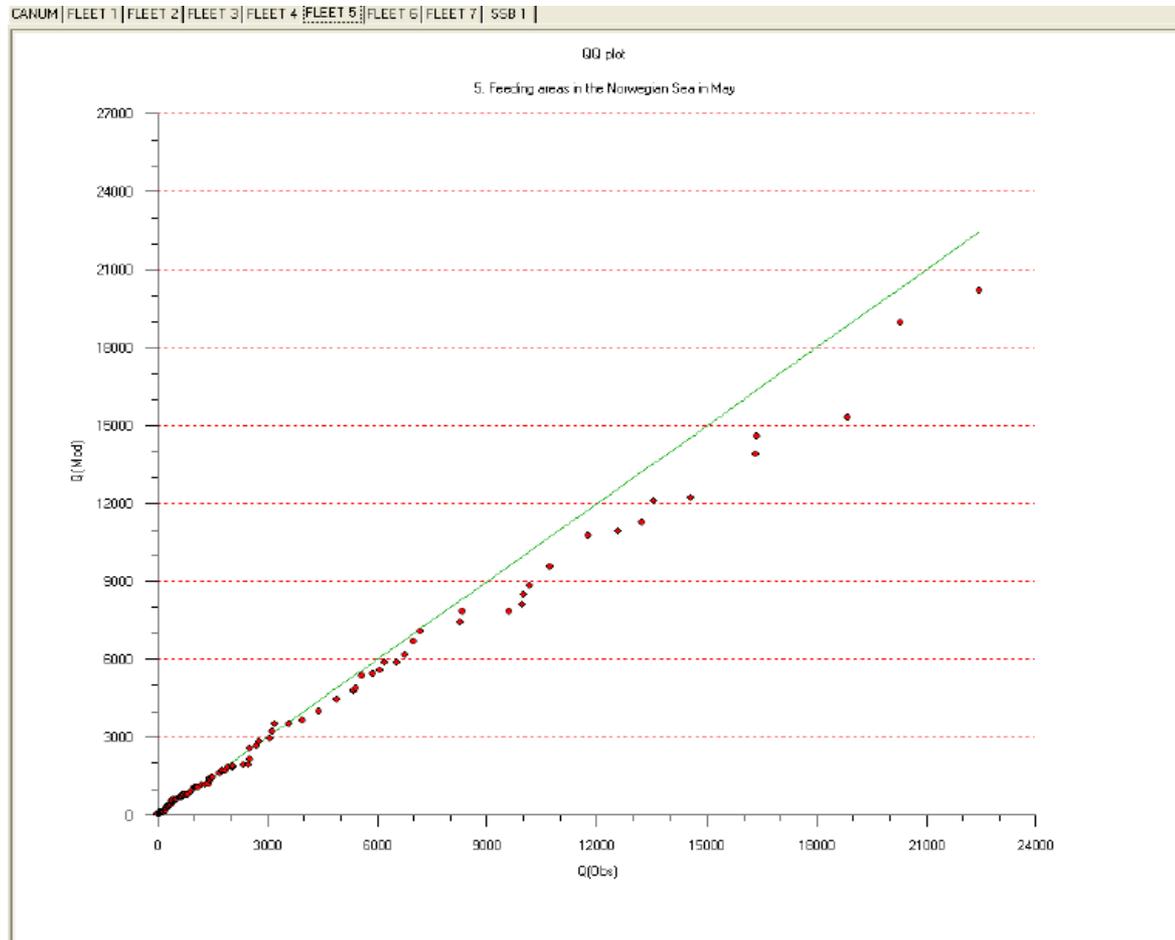
Полезен:

- Для проверки выбора целевой функции
- Для дальнейшего бутстрапа



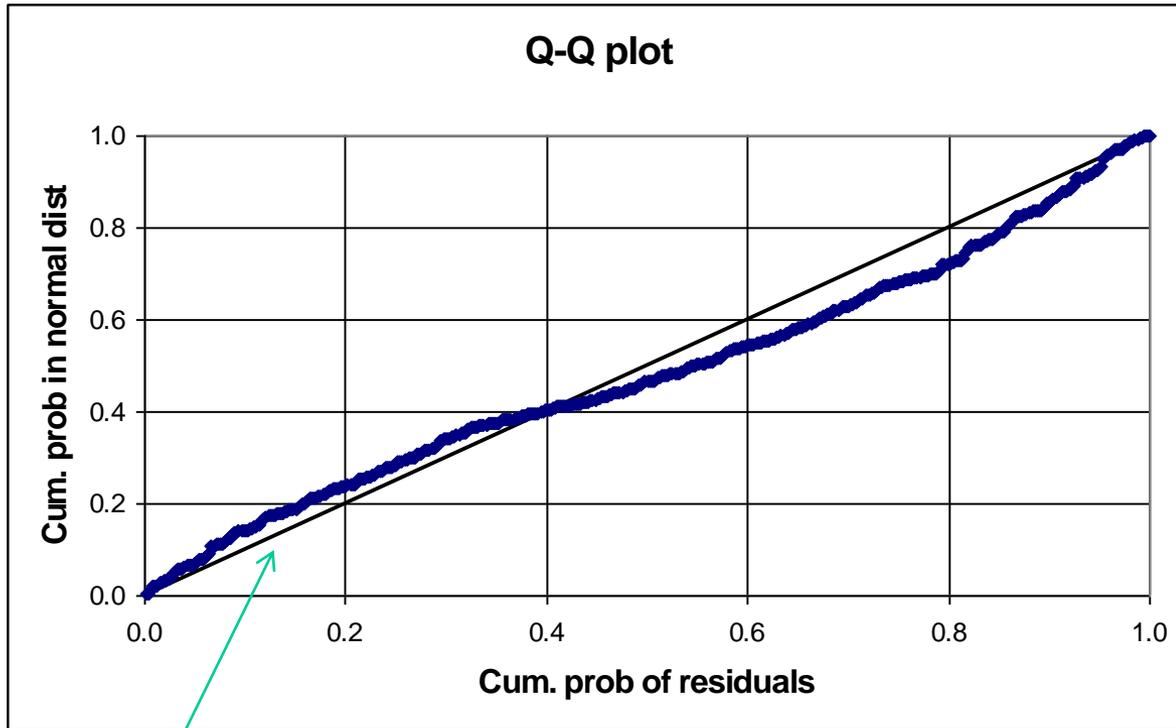
- 1- distribution of residuals in age composition of stock for real data
2- approximation by normal distribution

3. Анализ характера распределения остатков квантиль-квантильные графики (Q-Q plots)



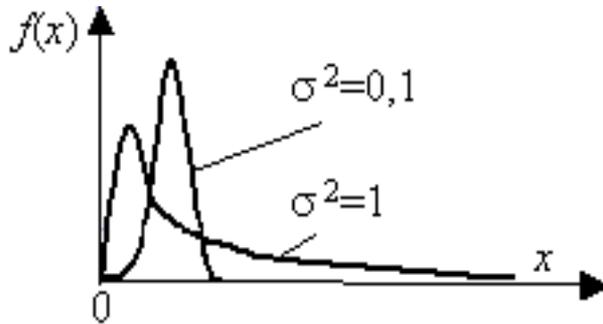
Тут в «реальном» распределении ошибок больше доля больших ошибок, чем в принятом распределении, т.е. можно попробовать распределение с более тяжёлыми хвостами

3. Анализ характера распределения остатков квантиль-квантильные графики (Q-Q plots)

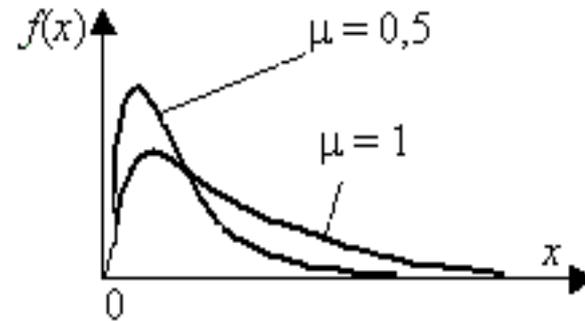


Здесь данные генерировались с гамма-распределенной ошибкой, а при оценке параметров использовалось предположение о логнормальном распределении ошибок. При правильном стандартном отклонении оно «широковато»

Многообразие встреченных на практике форм распределения наблюдений физических величин породило целые семейства их аналитической аппроксимации. Напомним лишь некоторые из них:

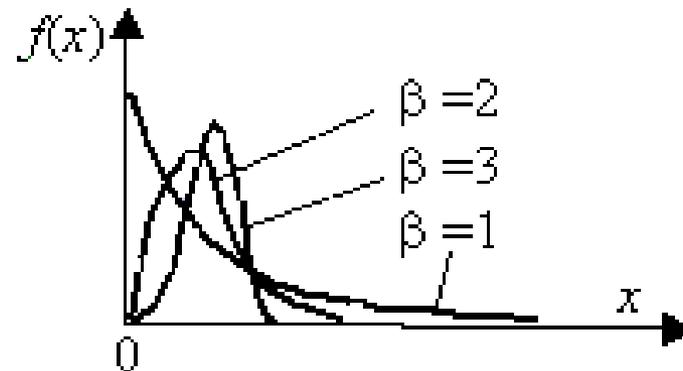
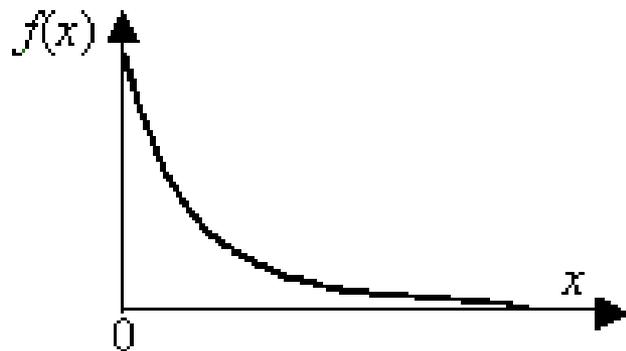


а)



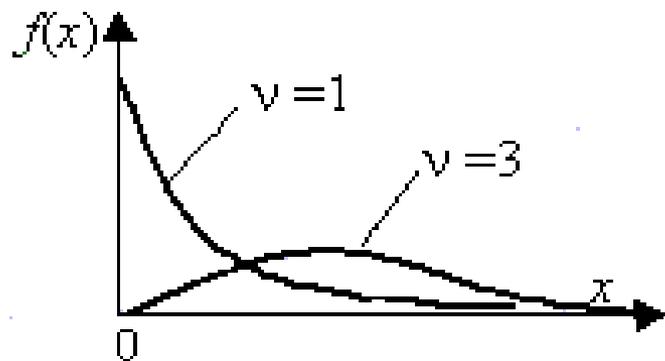
б)

Логарифмически нормальное: а) $\mu = 0$ б) $\sigma = 1$

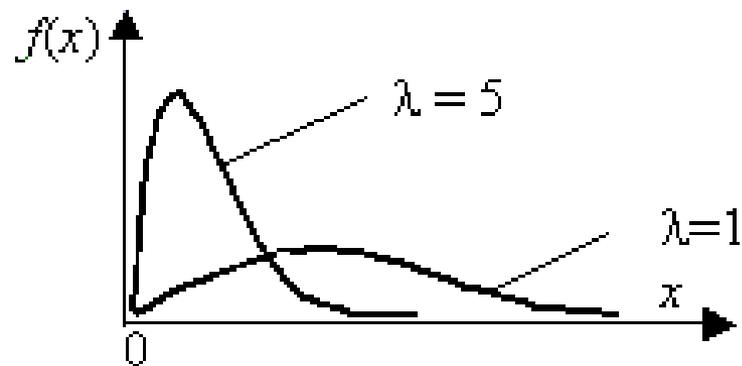


Экспоненциальное распределение

Распределение Вейбулла



а)



б)

Гамма-распределение: а) $\nu = 1$ б) $\nu = 3$

Казалось бы, их хватит, чтобы описать всё разнообразие жизни. Однако, как мы увидим позже, для работы с реальными ошибками и их оказалось мало.

Преимущество применения типовых законов распределения состоит в их хорошей изученности и возможности получения состоятельных, несмещенных и относительно высоко эффективных оценок параметров. Однако рассмотренные выше типовые законы распределения не обладают необходимым разнообразием форм, поэтому их применение не дает необходимой общности представления случайных величин, которые встречаются при исследовании систем

Функции плотности распределения для некоторых из распределений

Тип и функция плотности распределения	Математическое ожидание m_1 , дисперсия m_2 , асимметрия $b_1 = m_3/m_2^{3/2}$, эксцесс $b_2 = m_4/m_2^2$
Нормальное $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma^2}\right),$ $-\infty < x < \infty$	$m_1 = \mu_1, m_2 = \sigma^2 = 2,$ $b_1 = 0, b_2 = 3$
Логарифмически нормальное $\frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu_1)^2}{2\sigma^2}\right), x > 0$ $, 0, x \leq 0$	$m_1 = \exp(\mu_1 + 0,5\mu_2),$ $m_2 = \exp(2\mu_1 + \mu_2)(\exp(\mu_2) - 1),$ $b_1 = (\exp(\mu_2) + 2)\sqrt{\exp(\mu_2) - 1},$ $b_2 = \exp(4\mu_2) + 2\exp(3\mu_2) + 3\exp(2\mu_2) - 3$

Вейбулла

$$\frac{\beta}{\delta} \left(\frac{x}{\delta}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\delta}\right)^{\beta}\right), \quad x \geq 0,$$

$$0, \quad x < 0$$

$$\delta > 0, \beta > 0$$

$$m_1 = \delta g_1, \quad m_2 = \delta^2 (g_2 - g_1^2),$$

$$b_1 = (g_3 - 3g_1g_2 + 2g_1^3)/(g_2 - g_1^2)^{3/2},$$

$$a = (g_4 - 4g_1g_3 + 6g_2g_1^2 - 3g_1^4),$$

$$b_2 = a/(g_2 - g_1^2)^2,$$

$$g_i = \Gamma(1 + i/\beta)$$

Гамма

$$\frac{\lambda^{\nu}}{\Gamma(\nu)} x^{\nu-1} \exp(-\lambda x), \quad x \geq 0,$$

$$0, \quad x < 0$$

$$\nu > 0, \lambda > 0$$

$$m_1 = \nu/\lambda \quad m_2 = \nu/\lambda^2,$$

$$b_1 = 2/\sqrt{\nu},$$

$$b_2 = 3(\nu + 2)/\nu$$

Надо помнить, что:

классические функции правдоподобия исключительно не робастны

распространенные подходы:

- распределения с утяжеленными хвостами (в надежде лучше описать аутлаеры)
- смешанные (mixture) распределения
- экзотические сверхгибкие распределения, «натяжимые» на что угодно (но каков *смысл* этого? – масса параметров, при том натягиваемых на случайные аутлаеры)
- *квази-функции правдоподобия (quasi-likelihoods)*, основанные на снижении влияния «плохих» наблюдений (M-оценки)

Методы повышения робастности оценок:

-Цензурирование входных данных

(например, на основе статистических методов анализа, например, двумерных массивов данных (в т.ч. кригинг))

-Процедуры последовательного улучшения оценок, например, α -винзоризация

- Применение робастных статистик при минимизации

-Отход от классической статистики (нормальных распределений и метода наименьших квадратов)

- Логнормальная модель распределения ошибок

- Меры близости, свободные от вида распределения и изначально более устойчивые к резко выделяющимся значениям, чем сумма квадратов остатков

- сумма абсолютных остатков

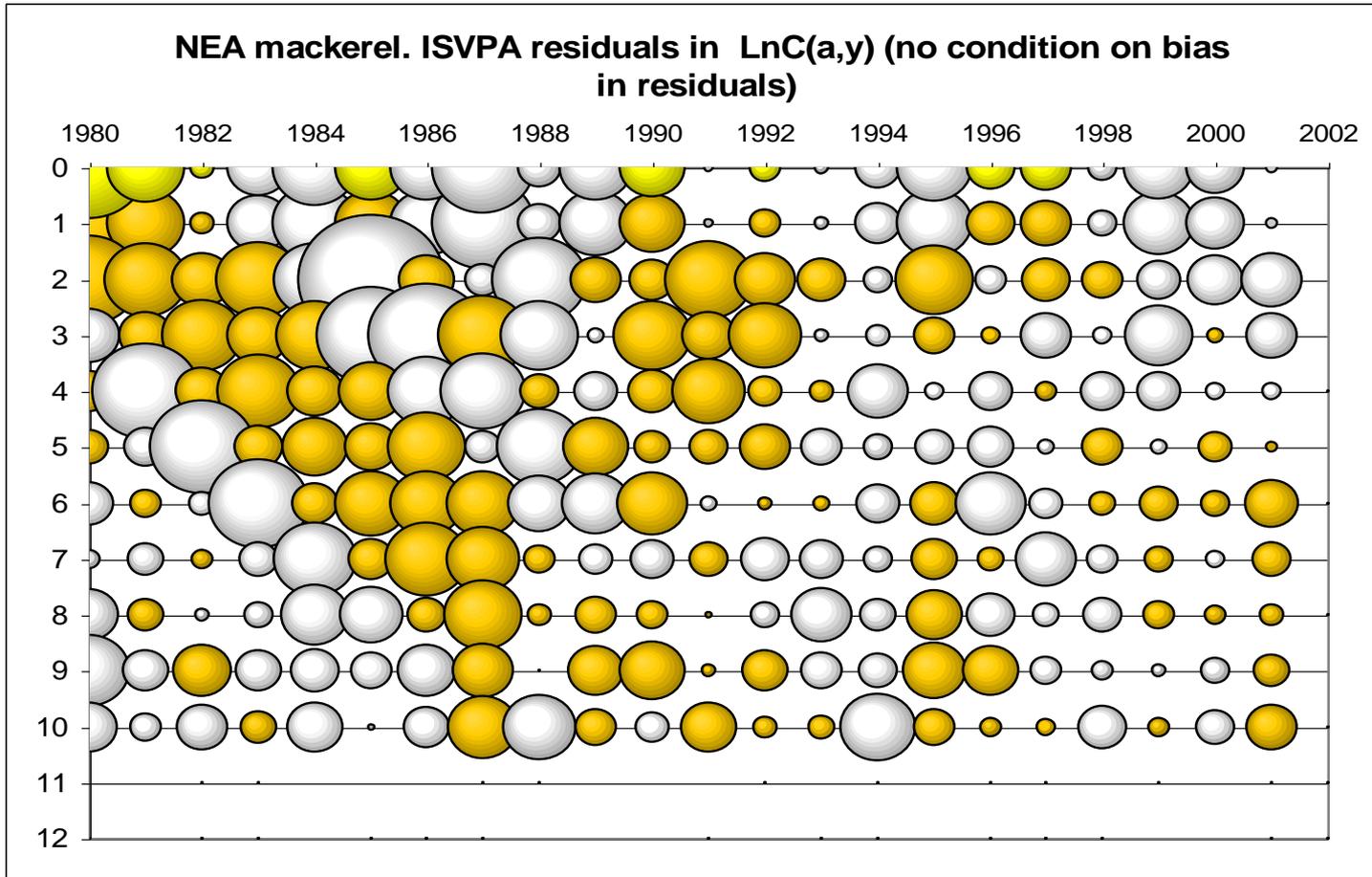
- α -урезанное среднее

- медиана распределения квадратных остатков

- абсолютное медианное отклонение (медиана распределения абсолютных отклонений остатков от медианы распределения остатков)

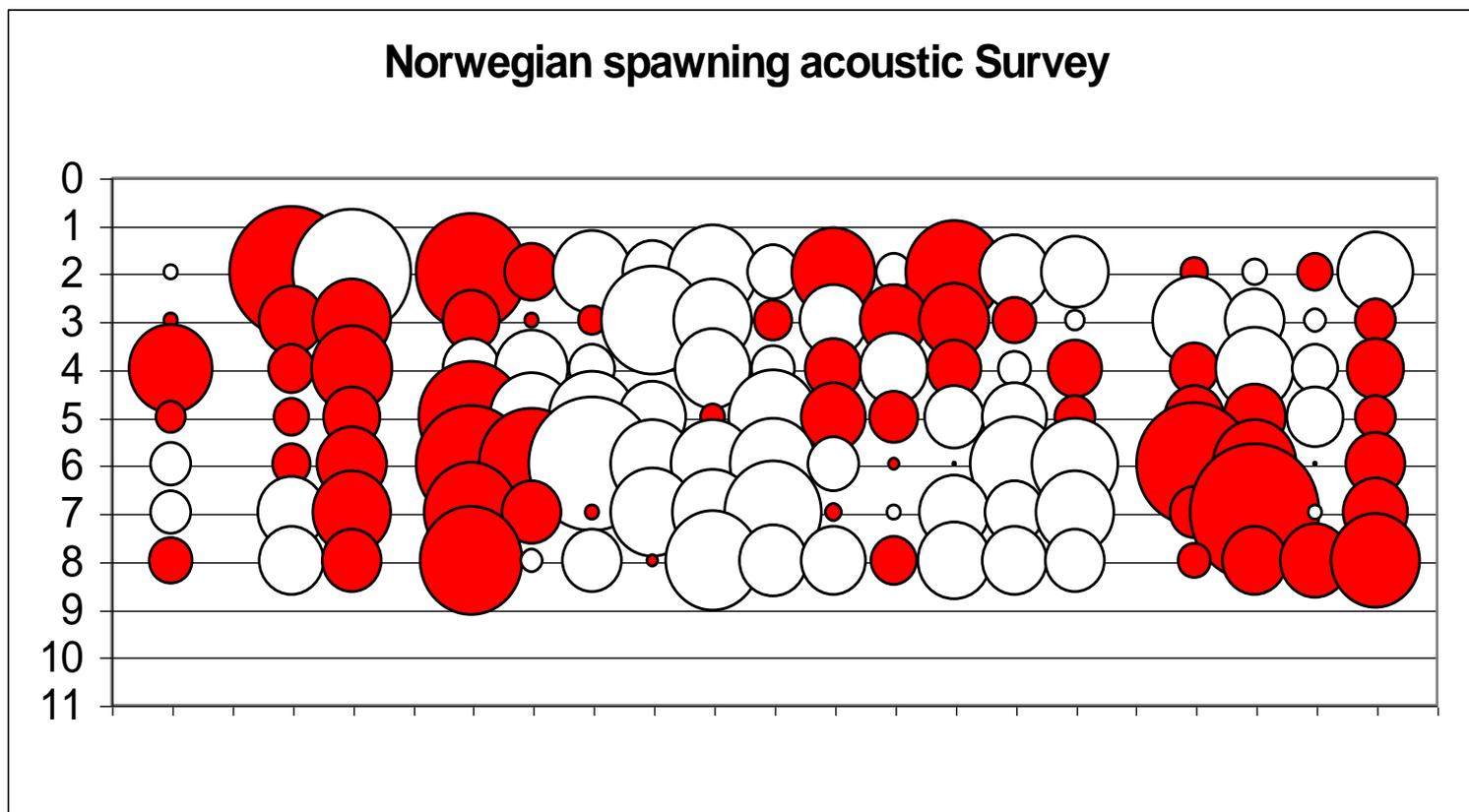
- оценка Ходжеса – Лемана (медиана распределения всех попарных средних, взятых из распределения остатков модели) и др.

3. Анализ структуры остатков



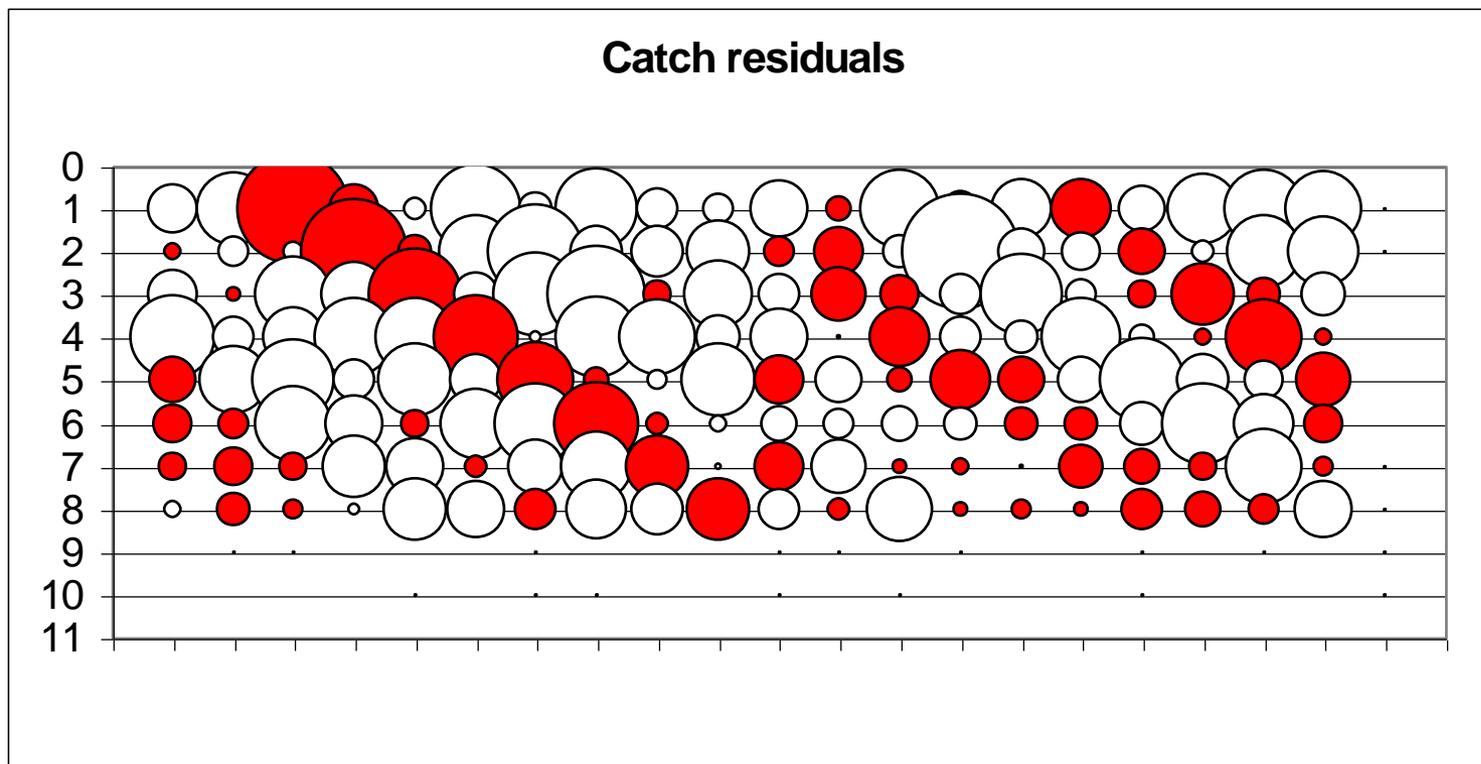
низкое качество данных для начального периода

3. анализ структуры остатков



«годовой эффект» - неоднородность условий съемок

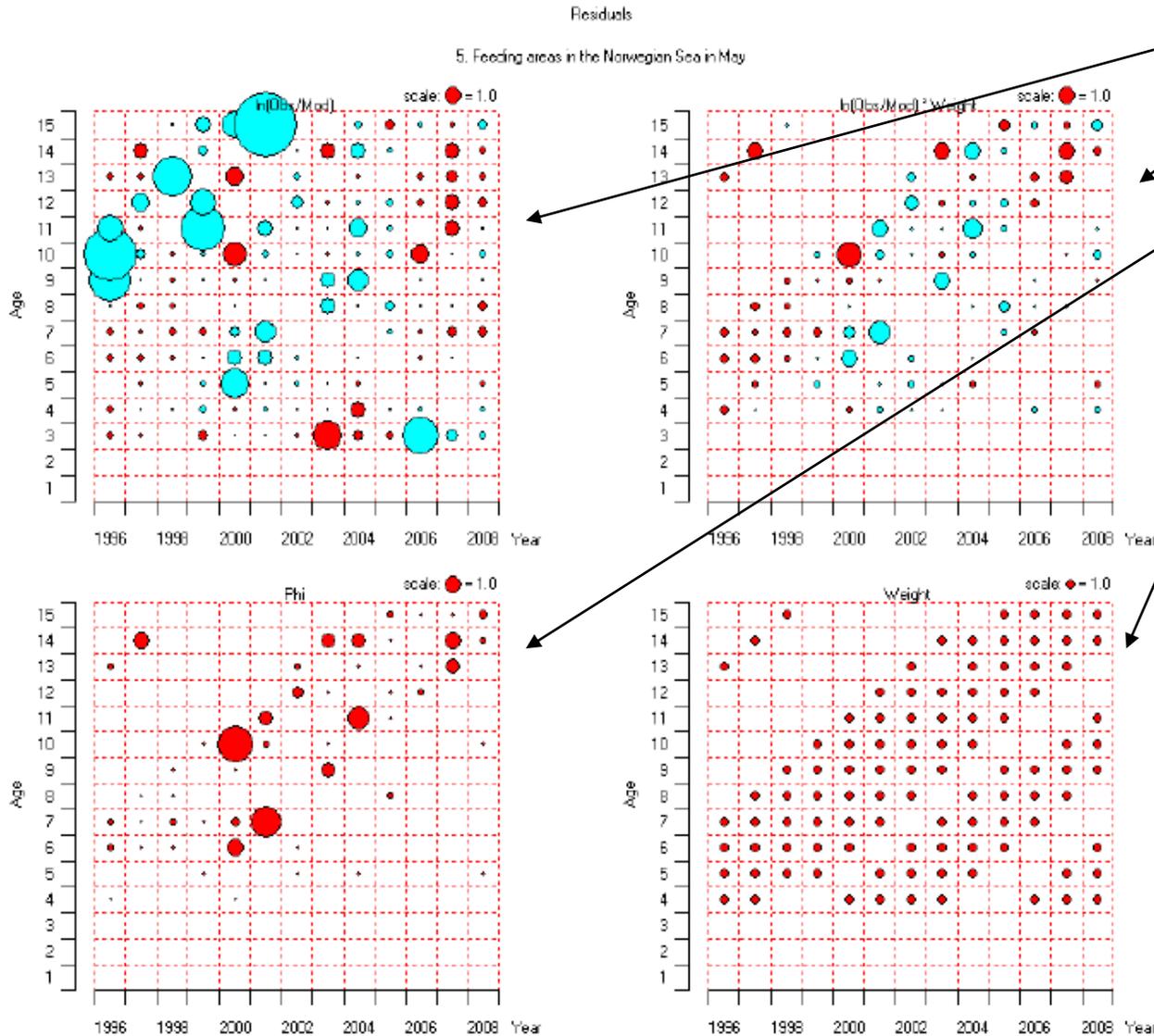
3. анализ структуры остатков



«когортный эффект» - проблемы с устойчивостью селективных свойств промысла и др.

Если использованы «поточечное» взвешивание данных (хотя с этим надо поступать осторожно), то удобно такое представление:

CANUM | FLEET 1 | FLEET 2 | FLEET 3 | FLEET 4 | FLEET 5 | FLEET 6 | FLEET 7 | SSB 1 |



Невзвеш. ост-ки

Взвеш. ост-ки

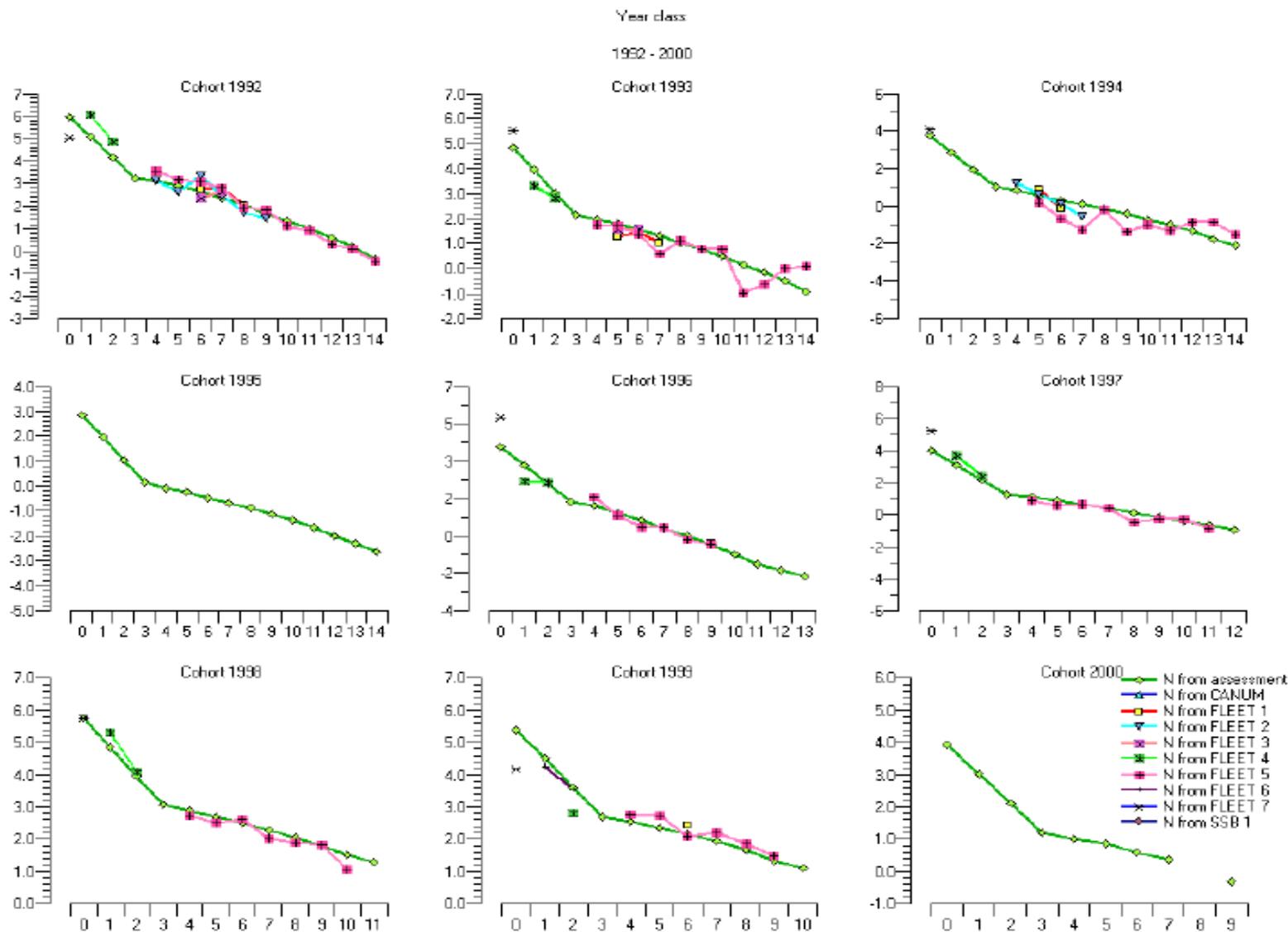
Вклад в целевую функцию

Веса

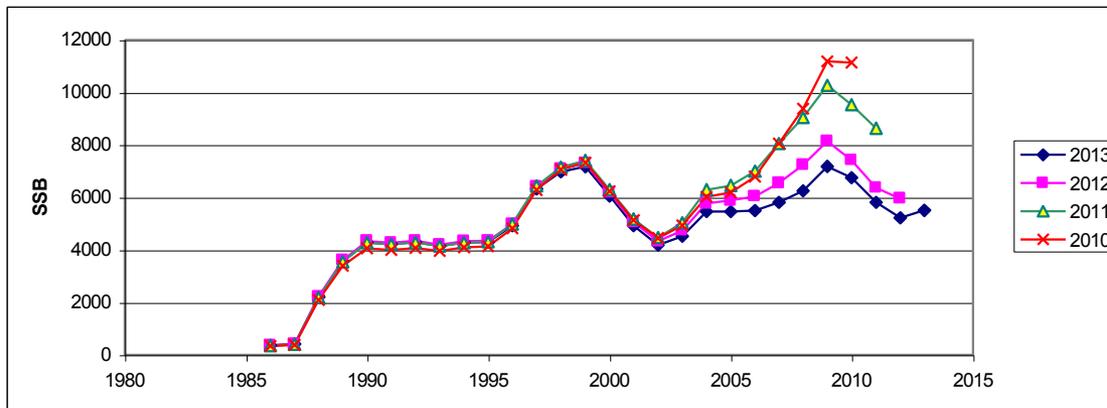
И так для каждого вида входных данных

Диагностика:

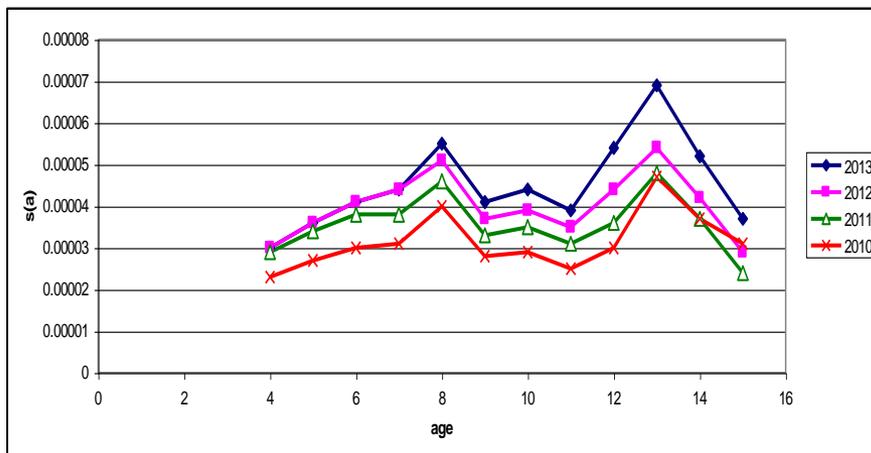
- То же, но устранены точки съемок, которым придан нулевой вес



Ретроспективный анализ (есть ли «историческое» смещение)



Одна из возможных причин – то, что принято за константу (например, q для съёмок), систематически меняется



например, из-за всё более «компактного» распределения запаса в результате присутствия в запасе всё меньшего числа разных возрастных классов при плохом пополнении за несколько последних лет

Оценка неопределенности – традиционные способы:

1. По ковариационной матрице для параметров θ_i .

Из определителя Гесса в рамках дельта-метода:

$$Cov_{ij} = 2 \Phi_{opt} / (m-n) * H_{ij}^{-1},$$

где компонент ii - дисперсия оценки параметра θ_i .

Φ_{opt} – оптимальное значение целевой функции Φ ,

H_{ij}^{-1} – компоненты обратной матрицы Гесса:

$$(H: H_{ij} = \partial^2 \Phi / \partial \theta_i \partial \theta_j)$$

$m-n$ – количество компонентов минус число параметров.

Определитель Гесса обычно получается в ходе квази-Ньютоновой оптимизации, но может быть посчитан и специально.

2. Путем бутстреппирования входных данных

Бутстреп -

Способ получить «что-то из ничего»



В одной из версий легенд о бароне Мюнхгаузене он вытянул себя не за волосы, а за петлю на ботфорте (bootstrap)

- *безусловный непараметрический*
(многократная перетасовка наблюдений)
- *условный непараметрический*
(многократная их генерация по распределению)
- *безусловный параметрический*
(добавление к наблюдениям перетасованных остатков)
- *условный параметрический*
(добавление остатков, сгенерир. по распределению)

Общая проблема

- “ Пусть есть набор наблюдений X_1, \dots, X_n , случайно выбранных из неизвестного распределения F . Нас интересует оценить некоторый параметр Θ , используя информацию по выборке, по которой есть оценка $\hat{\Theta} = t(x)$. Иметь некоторую оценку точности оценки также важно, как и саму оценку: нас интересует стандартная ошибка для $\hat{\Theta}$, а ещё лучше - доверительные интервалы для истинного значения Θ .”

– Эфрон и леПаж (1992)

Общая первичная гипотеза: наша выборка содержит всю информацию о генеральной совокупности, тогда

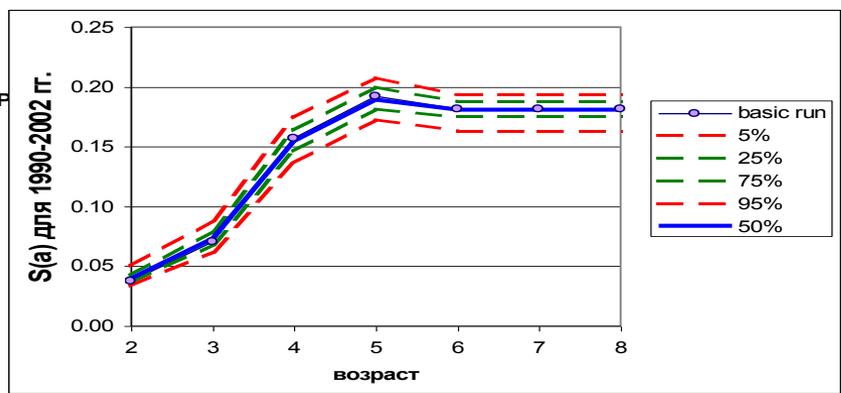
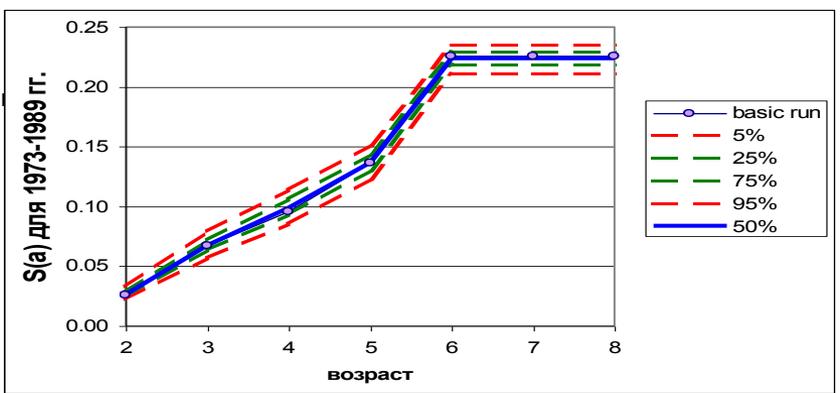
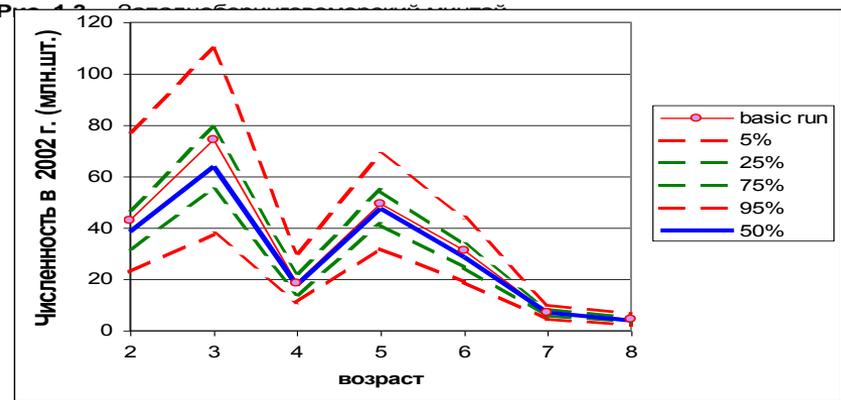
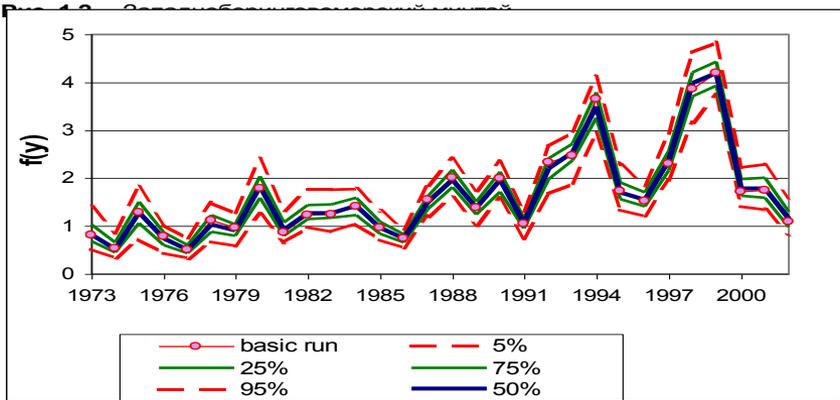
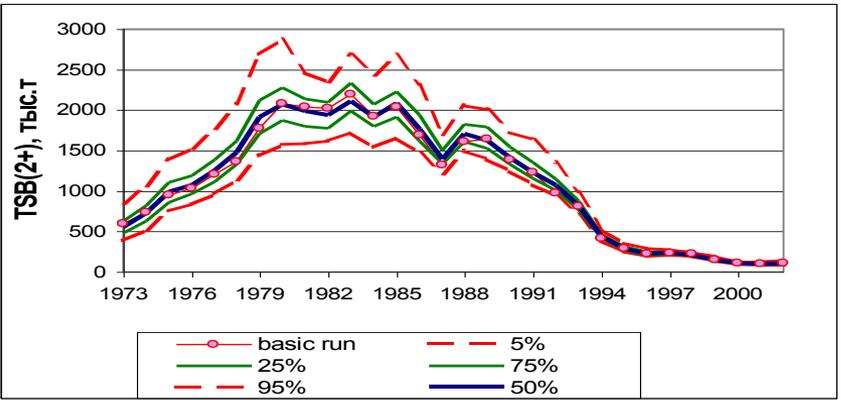
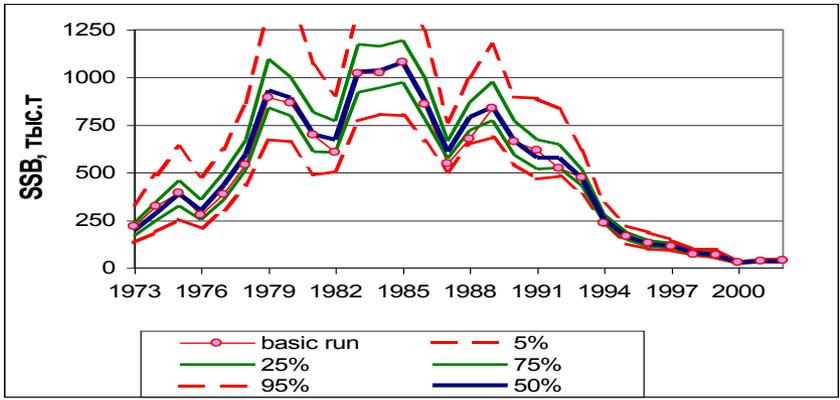
- Генерировать большое количество наборов псевдо-наблюдений $x_1, x_2, x_3, \dots, x_b$, из имеющихся
 - Наблюдения в каждом наборе данных генерируются путем случайного выбора из числа наблюдений в имеющемся наборе данных
 - Каждый набор псевдо-наблюдений имеет то же кол-во наблюдений, что и исходный набор.
- Прогон модели с каждым из этих наборов и подсчет интересующих статистик (стандартные отклонения, доверит. Интервалы, среднее, медиана и т.д.)
- Понятно, что этот простейший подход (безусловный параметрический) не всюду годится

Перед тем, как проводить бутстреп, надо задуматься:

- какой из видов бутстрепа уместен для тех или иных видов данных (catch-at-age, съемки с возр. структурой, и др.)

- что бутстрепировать: всю матрицу (ряд) с одними параметрами или со спецификой (например, данным для разных возрастных групп свойственна разная ошибка и т.д.; специфика отдельных лет с особо плохими данными – у них должна быть та же «привязка» или они могут кочевать и т.д.

Итак, провели бустреп



И задумались над результатами

- Например, увидели , что бутстреп-среднее сильно отличается от базового результата. Почему?
- А как себя ведет бутстреп-медиана? Тоже сильно отличается?
- А не взять ли другой вариант бутстрепа?
- А не взять ли другой закон распределения ошибок (если применен условный параметрический бутстреп) ?
- **А не взять ли вообще другой вариант самой оценочной модели и не начать ли всё заново?**

В рамках «непродуманного» бутстрепа порой могут генерироваться наборы псевдо-наблюдений «несовместимые с жизнью» самой модели или её процедуры оценки параметров:

Подумать, как это происходит и как с этим бороться