



## CALIF 4.0 ИНТЕРАКТИВНОЕ WEB-ПРИЛОЖЕНИЕ R SHINY ДЛЯ КАЛИБРОВКИ ВЕСОВ ОБСЛЕДОВАНИЯ

Семинар ЕЭК ООН по измерению бедности во время пандемии  
26 марта 2021 года

Борис Франкович  
Статистическое управление Словацкой Республики

# СОДЕРЖАНИЕ

- основы калибровки
- оптимальная стратегия
- Calif 4.0
- результаты словацкого обследования EU-SILC  
( [СДУЖ](#), Статистика доходов и условий жизни)



# МЕТОД КАЛИБРОВКИ

- население  $U$ , выборка  $S$
- каждая единица в  $S$  с расчетным весом  $d_k$  в качестве обратной вероятности отбора
- при возникновении отказа от ответов учитываются как ошибки репрезентативности - для этого необходимо подкорректировать веса
- без корректировки они не могут точно воспроизвести ни известные данные о численности населения  $X_j$ , ни расчетные характеристики



# МЕТОД КАЛИБРОВКИ

- в поисках новых весов  $w_k$  для каждой единицы в  $S$ , которые подтверждают  $X_j$  и минимально отличаются от начальных весов
- затем эти новые веса используются для оценки - они обрабатывают некоторый уровень смещения на отказ от ответов
- основан на методах оптимизации
- разница между исходным и калибровочным весом, вычисленная функциями расстояния с параметром  $r_k = \frac{w_k}{d_k}$
- могут быть установлены границы различий (некоторый допустимый интервал)



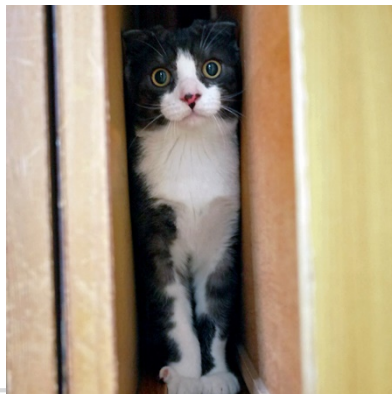
# МЕТОД КАЛИБРОВКИ

- предложен Девилем и Сарндалем
- значительно повышает точность подсчетов
- исследуемые переменные должны коррелировать со вспомогательными калибровочными переменными
- когда показатели ответов распределены неравномерно и это вызывает смещение
- вносит закономерность в обследования



# ОПТИМАЛЬНАЯ СТРАТЕГИЯ

- общее правило заключалось в том, чтобы найти такое решение, в котором границы были бы как можно более строгими
- это связано с восприятием крошечной разницы между исходным и калибровочным весом, когда они находятся в пределах некоторого напряженного интервала
- однако ограниченное пространство вынуждает исходные веса приближаться к границам, и в результате средняя разница в весе довольно высока



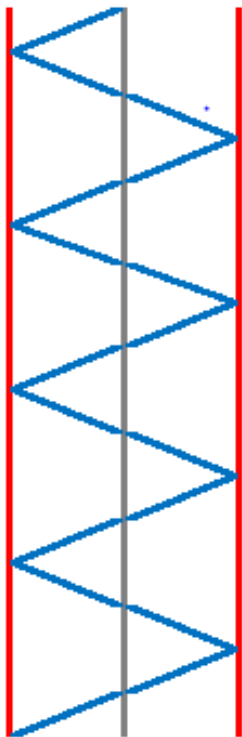
# ОПТИМАЛЬНАЯ СТРАТЕГИЯ

- менее строгие ограничения формируют некоторую область для перемещения, где лишь небольшое количество весов отодвигается к границам, в то время как другие остаются близкими к их исходным значениям
- таким образом, средняя разница между исходным и калибровочным весом минимальна для неограниченных методов
- чтобы избежать резкого колебания весов, следует использовать какой-нибудь ограничительный метод, но при этом должен быть найден реальный компромисс между шириной интервала и вносимыми смещениями



# ОПТИМАЛЬНАЯ СТРАТЕГИЯ

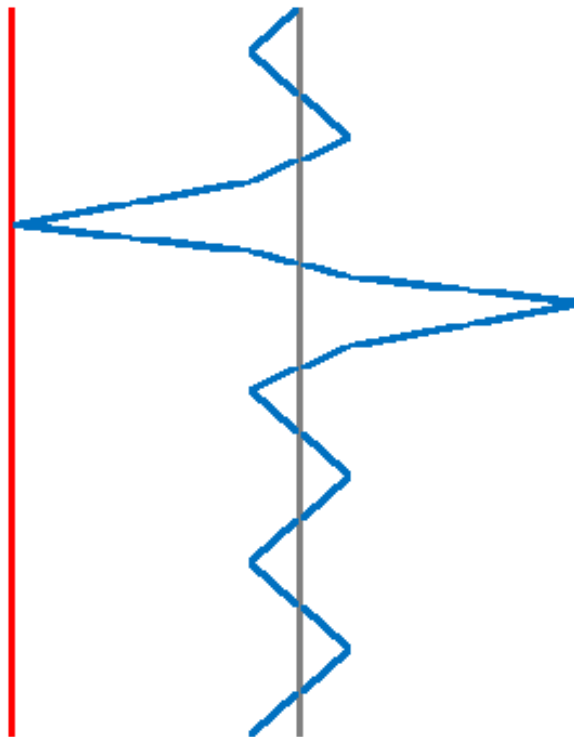
lower bound initial upper bound



lower bound

initial

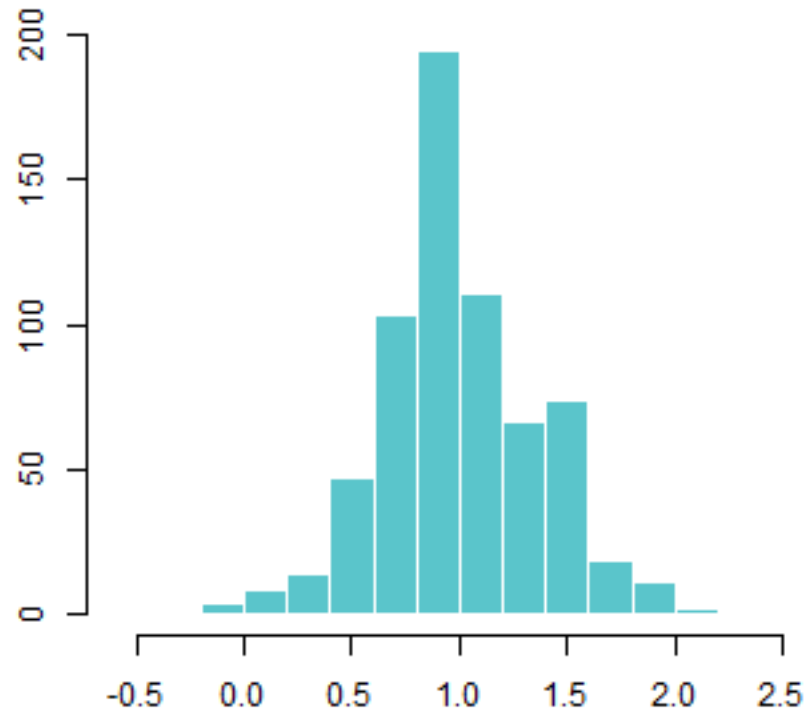
upper bound





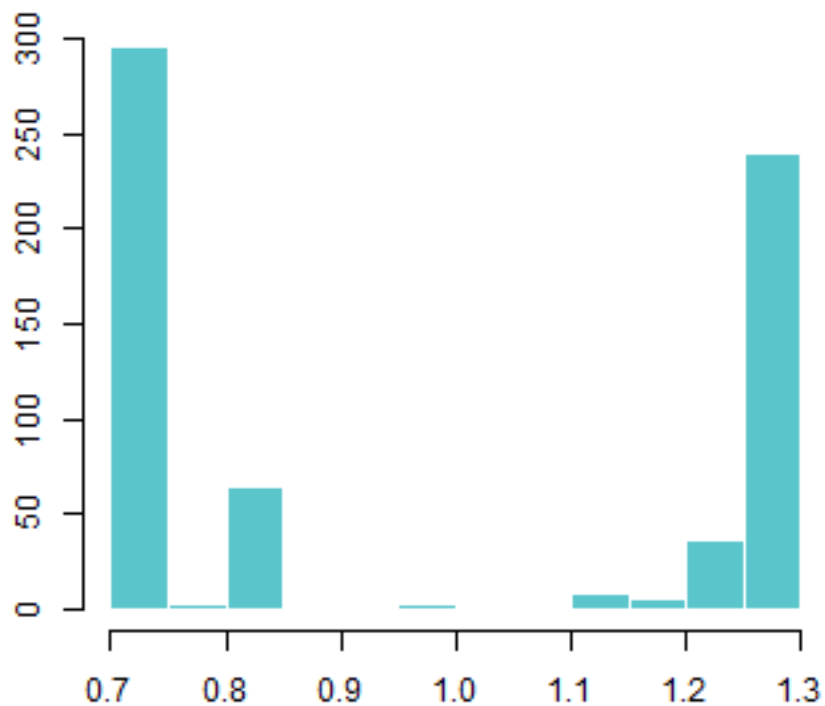
# ВЕСОВЫЕ КОЭФФИЦИЕНТЫ - БЕЗ ОГРАНИЧЕНИЙ

Histogram of quotients



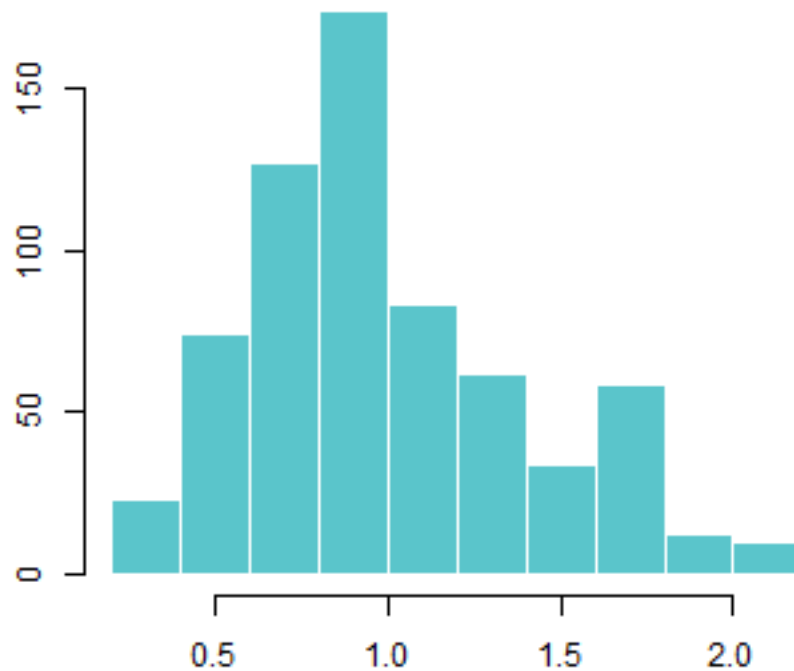
# ВЕСОВЫЕ КОЭФФИЦИЕНТЫ - СТРОГИЕ ОГРАНИЧЕНИЯ

Histogram of quotients



# ВЕСОВЫЕ КОЭФФИЦИЕНТЫ - СВОБОДНЫЕ ГРАНИЦЫ - ВОЗМОЖНЫЙ КОМПРОМИСС

Histogram of quotients



# ОПТИМАЛЬНАЯ СТРАТЕГИЯ - УСЛОВИЯ

- должны быть воспроизведены ограничения калибровки
- калибровочные веса не должны сильно отличаться от исходных весов
- они должны лежать в довольно напряженном интервале
- средняя разница весов должна быть небольшой по сравнению с линейным методом



# Calif 4.0

- веб-приложение R SHINY
- Лицензия GPL v3.0
- современный дизайн
- простой в работе, очень быстрый, надежный, проверенный временем метод
- веб-браузер выступает в качестве дисплея
- все вычисления выполняются локально в приложении R - данные в интернет не отправляются
- <https://github.com/SO-SR/Calif>
- <https://slovak.statistics.sk/wps/portal/ext/products/software.tools>



# ОСНОВНЫЕ ХАРАКТЕРИСТИКИ CALIF

- веб-дизайн GUI
- удобная эксплуатация
- двухступенчатая калибровка для интегрированных весов
- интерактивная загрузка данных
- стратификация
- подробные технические характеристики калибровочных переменных, определяемые пользователем
- функциональные и быстрые алгоритмы решения



# ОСНОВНЫЕ ХАРАКТЕРИСТИКИ CALIF

- широко используемые дистанционные функции
- возможны примерные решения
- создание закладок и интерактивное сохранение
- графические и полноценные выводимые данные
- и многое другое



# УСТАНОВКА

- установите R - бесплатно
- установите необходимые пакеты

```
install.packages(c('shiny', 'haven', 'sampling', 'nleqslv'))
```

- перейдите на веб-страницу github или SO SR's и выполните соответствующую команду

```
shiny::runGitHub('Calif', 'SO-SR', destdir = getwd(),  
  launch.browser = TRUE)
```

- или

```
shiny::runUrl('https://slovak.statistics.sk/wps/wcm/connect/7014  
bfd4-54a2-4080-929f-eb949bf25e39/calif.zip?MOD=  
AJPERES&CVID=m7Xjumj&CVID=m7Xjumj', filetype = '.zip',  
  destdir = getwd(), launch.browser = TRUE)
```





# ОБЗОРНАЯ ВКЛАДКА:

Calif 4.0 Overview Data Calibration

## Welcome to Calif!

Calif is the Shiny web application used for calibration of weights in statistical surveys. It is an open-source software and you are free to use it, also without any knowledge of R programming language. Detailed information on how to use Calif can be found in the [Manual](#)

### Getting started

To get started working with Calif, you first need to prepare the data and the table of totals (which are the known population marginals). You can learn it in the Manual, in fact it is very easy and intuitive. Once they are ready, you need to load them into Calif in the [next tab](#). After that, you select the calibration variables, either categorical or numerical, initial weights for calibration, possible stratification and calibration methods.

### Optimal strategy

Then you can calibrate the weights. In Calif, after each calibration step, you are provided with useful statistics that can help you to decide which method and parameters are most suitable for performing the best possible calibration. The optimal strategy is to find the new weights that reproduce the population totals and are as close as possible to the initial weights such that the lower and upper bounds form a narrow interval (optimally close to 1 from both sides) and the average difference between the initial and calibration weights is low in comparison to the linear solution. The feasibility of the average difference is demonstrated by a pie chart on the **Calibration** tab. If you set too tense bounds, even if the solution is found, the histogram of quotients on the **Calibration** tab will look unnaturally, the average difference will be high and such calibration will not be appropriate.

### Working directory

Your actual working directory is set to `D:/Users/Frank2/Desktop`. If you want to change it, you can do it by clicking on this [button](#)

### Bookmarking

You can bookmark your work at any time by clicking on the **Bookmark** button on the **Calibration** tab. The files and settings are saved into your working directory and you obtain a link that can be used to reproduce your work when Calif is run again.

### Output

You can save your work at any time by clicking on the **Save** button on the **Calibration** tab. The output is an identical copy of your data with column `CalibrationWeight` added. This column represents your latest calibration attempt, i.e. if you tried several calibrations with different settings, the last one is recorded. If you haven't performed any calibration yet, `CalibrationWeight` column contains the initial weights. If you carry out stratified calibration, `CalibrationWeight` column contains the new weights for strata that have been already calibrated and the initial weights for strata that have remained untouched. Although the latest state of play is still remembered throughout all strata, you are recommended to regularly save or bookmark your work in order not to lose it if some unexpected error occurs.



# Вкладка «Данные»

Period  Comma

Specify calibration variables

Stratification

Weights:

Stratification variable:

Numerical variables: s1a6, s2a1, s2a2, s2a3, s2a4, s2a5, s2a6, MEMBERS, Weight

Categorical variables: s1a6, s2a1, s2a2, s2a3, s2a4, s2a5, s2a6, MEMBERS, Weight

Other settings

Indicators monitoring (optional): id\_hd, REGION, s1a1, s1a2, s1a3, ...

Total (%)

Total (value)

Mean (%)

Mean (value)

Missings:

Eliminate rows with missings

Maximum nr of iterations:

Tolerance:

Dataset with 6 rows and 18 columns

Show  entries

Search:

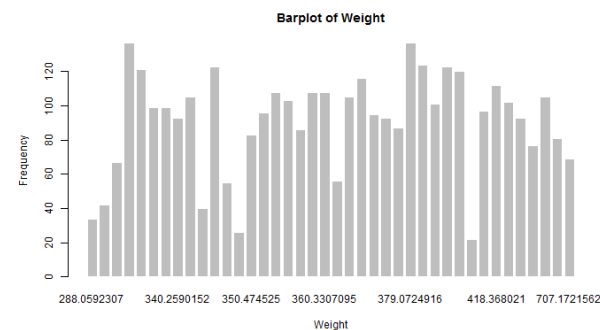
STRA	MEMBERS_1	MEMBERS_2	MEMBERS_3	MEMBERS_4	MEMBERS_5	s1a1	s1a2	s1a3	s1a4	s1a5	s1a6	s2a1	s2a2	s2a3	s2a4	s2a5	s2a6
1	64841	58791	48099	48019	15489	45907	34665	99262	39020	38788	29654	43365	33689	101594	45010	472	
3	50982	48547	40169	47186	26081	43774	42256	94469	43697	37300	29888	41784	40310	89073	43526	407	
4	70510	59378	48030	57141	26692	52052	47017	109497	50025	43107	32739	48754	44584	104615	50507	492	
5	55249	53313	42573	49341	37621	60506	52184	109673	48127	39322	28641	57549	50221	103446	47445	436	
6	64808	59340	47747	53156	21460	53872	45809	103172	47450	39862	29400	51399	44327	99249	48260	452	

Showing 1 to 5 of 6 entries

Previous   Next

## Explore variables

Explore as categorical

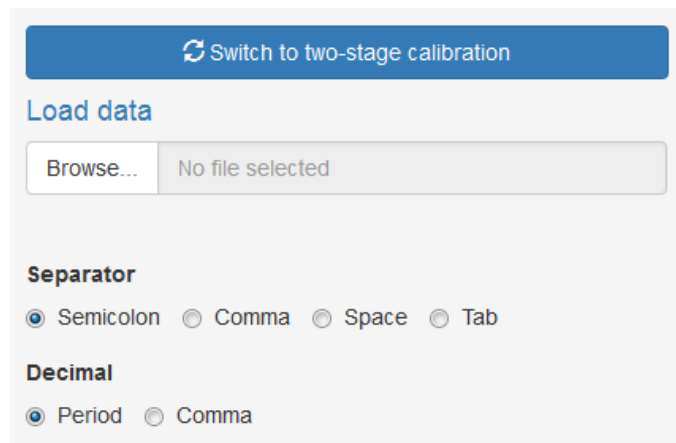


Value	Frequency
288.0592307	34
300.0748022	42
317.0452195	67
327.6444278	137
334.3032222	121
339.8685987	99
339.9882421	99
340.2590152	93
340.7154621	105
345.0188148	40
346.7000684	123



## Вкладка «Данные»

- импорт данных и таблица общих показателей
- Поддерживаются форматы .csv, .txt и .sas7bdat.
- данные отображаются в интерактивном режиме
- переключение между одноступенчатой и двухступенчатой калибровкой



Switch to two-stage calibration

Load data

Browse... No file selected

**Separator**

Semicolon  Comma  Space  Tab

**Decimal**

Period  Comma



## Вкладка «Данные»

- структура данных может быть практически произвольной - она может содержать любые переменные, но как минимум калибровочные переменные и начальные веса
- необходимая структура таблицы общих показателей описана в "Руководстве"
- после калибровки возвращаются данные с одинаковой структурой с добавлением одной колонки



# ВКЛАДКА «КАЛИБРОВКА»

Calif 4.0 Overview Data Calibration

**Choose strata**

1  
3  
4  
5  
6  
2  
7  
8

Show with initial weights

**Method & Solver**

**Method** **Solver**

Linear  calib

Raking ratio  nleqslv

Logit

Linear bounded

**Lower bound** **Upper bound**

0,3 2

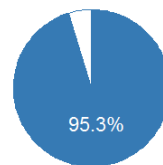
**CALIBRATE**

Bookmark Save

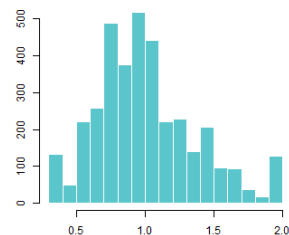
## Results

Initial weights interval	288.059	707.172
Calibration weights interval	86.418	1260.294
Lower bound obtained	0.300	
Upper bound obtained	2.000	
Average weight quotient	1.008	
Average difference	105.647	
Minimum realistic lower bound	0.004	

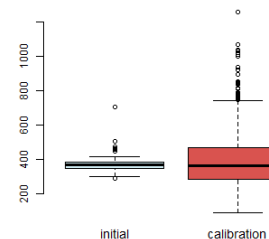
## Average difference feasibility



## Histogram of quotients



## Boxplots of weights



## Totals obtained

Show obtained totals as values

Stratum	MEMBERS_1	MEMBERS_2	MEMBERS_3	MEMBERS_4	MEMBERS_5	s1a1	s1a2	s1a3	s1a4	s1a5	s1a6
1	100.00%	100.00%	100.00%	100.00%	100.00%	97.88%	100.00%	100.00%	100.00%	100.00%	100.00%
3	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
4	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
6	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

## Weights & Quotients

Show 20 entries

Row	Initial	Calibration	Quotients
3266	288.05923	86.41777	0.3



## ВКЛАДКА «КАЛИБРОВКА»

- выбор группы, подходящего метода и алгоритма решения
- отображение с начальными весами - Н-Т оценки калиброванных итогов с использованием начальных весов по сравнению с этими общими показателями - разница между фактической реальностью и желаемой целью - очень полезный инструмент для изучения распределения отсутствующих ответов в обследовании - значения на большом расстоянии от 100% подвержены погрешностям при отсутствии выборки из-за отсутствия ответов
- *калибр + линейно ограниченный* идеально подходит для социальных обследований (например, EU-SILC)



# ВКЛАДКА «КАЛИБРОВКА»

## Method & Solver

### Method

- Linear
- Raking ratio
- Logit
- Linear bounded

### Solver

- calib
- nleqslv

### Lower bound

0.5

### Upper bound

2



# ВКЛАДКА «КАЛИБРОВКА»

- сводная статистика - получены общие показатели, первый и последний интервалы веса, достигнуты весовые коэффициенты, средняя разница ( $AD = \frac{1}{n} \sum_{k=1}^n |w_k - d_k|$ )

## Results

Initial weights interval	286.604	337.132
Calibration weights interval	85.981	761.426
Lower bound obtained	0.300	
Upper bound obtained	2.657	
Average weight quotient	1.000	
Average difference	98.330	
Minimum realistic lower bound	0.004	

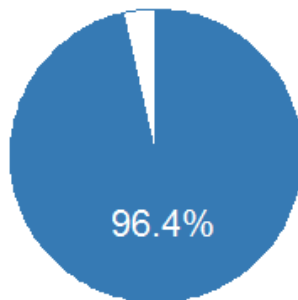




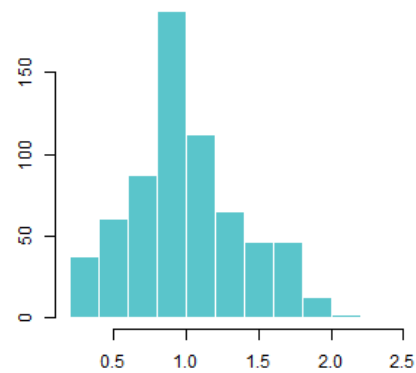
# ВКЛАДКА «КАЛИБРОВКА»

- полезные цифры

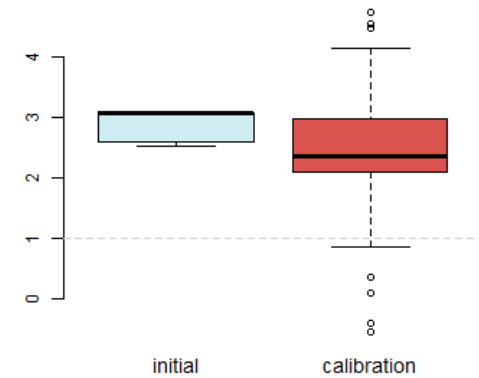
Average difference feasibility



Histogram of quotients



Boxplots of weights



# РЕЗУЛЬТАТЫ EU-SILC 2016

- каждая новая панель калибруется отдельно для всей популяции
- Новая панель 2016 г. - 1510 домохозяйств
- критерии калибровки отдельно на каждом уровне NUTS3
  - 8 половозрастных групп
  - работники
  - безработные
  - пенсионеры
  - 2 категории размера домохозяйств



# РЕЗУЛЬТАТЫ EU-SILC 2016

- всего 104 общих показателя для калибровки
- итоговые результаты на 100% воспроизводятся с дополнительными настройками

NUTS3	Алгоритм решения	Метод	Нижний предел	Верхний предел
1	Калибр.	линейно ограниченный	0,3	3
2	Калибр.	линейно ограниченный	0,3	2,7
3	Калибр.	линейно ограниченный	0,3	2,2
4	Калибр.	линейно ограниченный	0,3	2,2
5	Калибр.	линейно ограниченный	0,4	2,2
6	Калибр.	линейно ограниченный	0,2	2,8
7	Калибр.	линейно ограниченный	0,3	2,5
8	Калибр.	линейно ограниченный	0,3	2,2



# РЕЗУЛЬТАТЫ EU-SILC 2016

- весь перекрестный файл откалиброван еще раз
- перекрестный файл 2016 г. - 5738 домохозяйств
- критерии калибровки отдельно на каждом уровне NUTS3
  - 12 половозрастных групп
  - работники
  - индивидуальные предприниматели
  - безработные
  - пенсионеры
  - 2 категории размера домохозяйств



# РЕЗУЛЬТАТЫ EU-SILC 2016

- всего 144 общих показателя для калибровки
- итоговые результаты на 100% воспроизводятся с дополнительными настройками

NUTS3	Алгоритм решения	Метод	Нижний предел	Верхний предел
1	Калибр.	линейно ограниченный	0,5	1,5
2	Калибр.	линейно ограниченный	0,5	1,4
3	Калибр.	линейно ограниченный	0,5	1,3
4	Калибр.	линейно ограниченный	0,5	1,5
5	Калибр.	линейно ограниченный	0,5	1,3
6	Калибр.	линейно ограниченный	0,5	1,3
7	Калибр.	линейно ограниченный	0,5	1,4
8	Калибр.	линейно ограниченный	0,5	1,3



СПАСИБО ЗА ВНИМАНИЕ!



STATISTICAL  
OFFICE  
OF THE SLOVAK  
REPUBLIC