

# **ТИПОЛОГИЧЕСКАЯ ГРУППИРОВКА НА ОСНОВЕ ДЕКОМПОЗИЦИИ СМЕСЕЙ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ**

**С.Е. Хрущев, Ю.Н. Исмайлова**

**15 июня 2022**







# Задача

Рассмотрим некую неоднородную совокупность (популяцию), которая состоит из  $k$  подгрупп (субпопуляций). Предположим, что наблюдаемый признак (характеристика) внутри  $i$ -ой подгруппы распределен в соответствии с функцией распределения  $F_i(t)$ , а вероятность того, что случайно выбранный индивид из всей популяции будет именно из  $i$ -ой субпопуляции равна  $p_i$ . Тогда вероятность того, что у случайно выбранного из всей совокупности индивида значение наблюдаемого признака будет меньше, чем  $t$ , равна

$$F(t) = p_1 F_1(t) + \dots + p_k F_k(t).$$

Будем предполагать, что нам необходимо определить тип субпопуляции, к которому принадлежит наблюдение, т.е. требуется решить обратную операцию смешивания, а именно операцию разделения (декомпозиции) смеси. И, как следствие, получим типологию рассматриваемой неоднородной популяции.

# Смесь вероятностных распределений

Пусть исследуется неоднородная выборка  $X = (X_1, X_2, \dots, X_n)$ , взятая из конечной смеси вероятностных распределений, т.е.

$$F(t) = p_1 F_1(t) + \dots + p_k F_k(t),$$

где веса  $p_i \geq 0$ ,  $i = 1, \dots, k$  – смешивающие вероятности, удовлетворяющие условию  $p_1 + \dots + p_k = 1$ ,

$F_i(t)$ ,  $i = 1, \dots, k$  – функции распределения (компоненты смеси), соответствующие смешиваемым вероятностным распределениям. Будем предполагать унимодальность компонент смеси.

Если компоненты смеси абсолютно непрерывны и имеют плотность  $f_i$ , то смесь также будет абсолютно непрерывной с плотностью

$$f(t) = p_1 f_1(t) + \dots + p_k f_k(t).$$

## Теоретическая задача

Пусть компоненты смеси  $f_i$  определяются неким известным параметрическим семейством, т.е.  $f_i(t) \equiv f_i(t, \theta_i)$ , где  $\theta_i$  - неизвестный параметр (скалярный или векторный). Имеем

$$f(t) = p_1 f_1(t, \theta_1) + \dots + p_k f_k(t, \theta_k).$$

Задачей разделения (декомпозицией) смеси называют задачу нахождения неизвестных параметров  $\vec{\theta} = (p_1, \dots, p_k, \theta_1, \dots, \theta_k)$ .

**Пример.** Пусть  $f \sim U_{0,1}$ , и компоненты смеси имеют равномерное распределение. Тогда справедливо, что

$$\begin{aligned} I(t \in (0; 1]) &= 0,5 * 2I(t \in (0; 0,5]) + 0,5 * 2I(t \in (0,5; 1]) \\ &= 0,25 * 4I(t \in (0; 0,25]) + 0,25 * 4I(t \in (0,25; 0,5]) + 0,5 * 2I(t \in (0,5; 1)). \end{aligned}$$

# Идентифицируемые конечные смеси

**Опр.** Конечная смесь называется идентифицируемой, если из равенства

$$\sum_{i=1}^k p_i f_i(t, \theta_i) = \sum_{j=1}^m q_j f_j(t, \theta_j)$$

вытекает, что  $k = m$  и для каждого индекса  $i$  существует индекс  $j$  такой, что  $p_i = q_j$ ,  $f_i(t, \theta_i) \equiv f_j(t, \theta_j)$ .

**Замечание.** Идентифицируемыми являются конечные смеси из распределений: нормальных, гамма, Коши, пуассоновских, биномиальных (с известной вероятностью успеха).

## Практическая задача

Пусть компоненты смеси  $f_i$  определяются неким известным параметрическим семейством, т.е.  $f_i(t) \equiv f_i(t, \theta_i)$ , где  $\theta_i$  - неизвестный параметр (скалярный или векторный). Имеем

$$f(t) = p_1 f_1(t, \theta_1) + \dots + p_k f_k(t, \theta_k).$$

**Задачей разделения (декомпозицией)** смеси называют задачу статистического оценивания неизвестных параметров  $\vec{\theta} = (p_1, \dots, p_k, \theta_1, \dots, \theta_k)$  по известной реализации выборки. При этом предполагается, что смесь является идентифицируемой.



## Как искать оценки?

**Метод моментов** (Pearson K., 1894): в частных случаях позволяет найти явный вид оценок. В общем случае оценки искать сложно. Оценки хуже, чем в ММП.

**Метод максимального правдоподобия.** Оценки лучше, но в явном виде не найти (только в некоторых частных случаях). Оценки ищут с помощью специальных итерационных методов. Отметим, что ОМП может и не существовать.

**Пример.** Рассмотрим двухкомпонентную смесь с равными весами произвольного нормального и стандартного нормального распределений:

$$f(t) = 0,5 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-a)^2}{2\sigma^2}} + 0,5 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

$$\psi(x, a, \sigma^2) = \prod_{i=1}^n \left( 0,5 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-a)^2}{2\sigma^2}} + 0,5 \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \right),$$

$$\psi(x, a = x_1, \sigma^2) \rightarrow \infty \text{ при } \sigma^2 \rightarrow 0.$$

## **ЕМ-алгоритм, SEM-алгоритм**

**ЕМ-алгоритм** (Dempster, Laird and Rubin, 1977): не очень быстр, не гарантирует сходимость к глобальному максимуму, обладает неустойчивостью по начальным данным.

**SEM-алгоритм** (Celeux G. and Diebolt J., 1985 г.) (+ модификации (Королев В.Ю. с др., Горшенин А.К.)): быстр, обладает устойчивостью к начальным данным; как правило, сходится к глобальному максимуму.

## Определение числа компонент смеси.

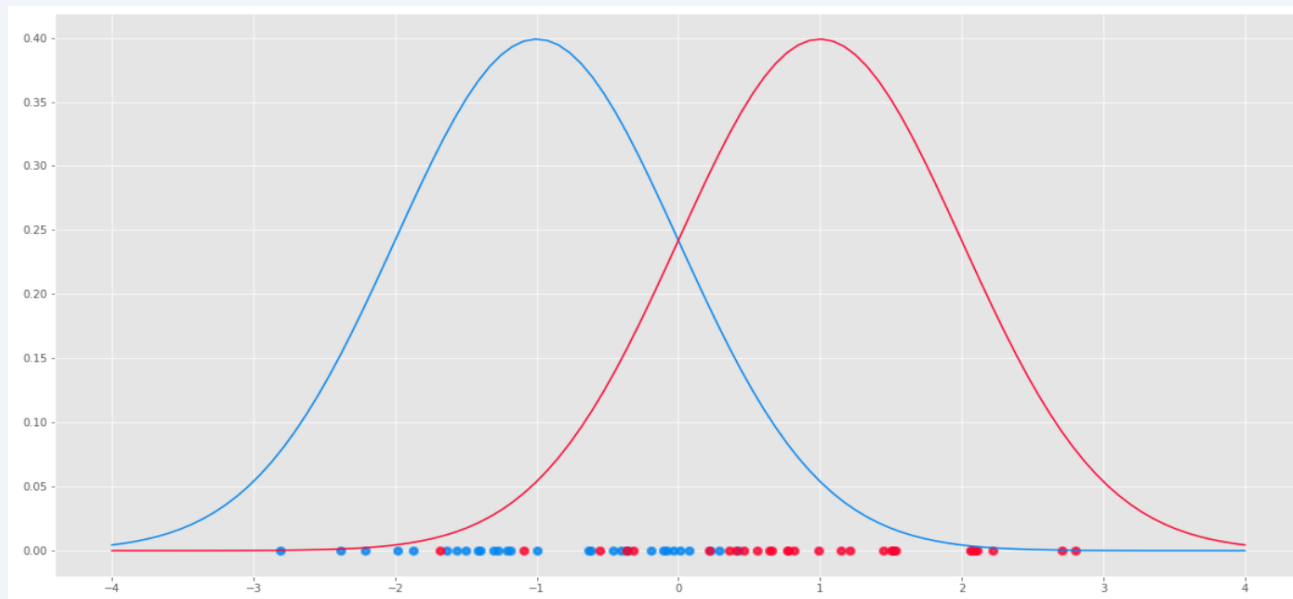
На практике параметр  $k$  (число компонент смеси), может быть неизвестен. Выбор слишком малых значений для него в большинстве случаев ухудшает качество аппроксимации распределения данных. Неоправданное увеличение числа компонент хоть и будет увеличивать правдоподобие, но приведет к искажению интерпретации результатов.

Многие существующие подходы к определению числа компонент смеси основываются на понятии расстояния Кульбака-Лейблера и носят название информационных. Для корректности применения таких информационных критериев требуется выполнение условий регулярности (в частности, конечность функции правдоподобия).

$AIC = -2 \ln \psi(x; \hat{\theta}(x)) + 2d$ ,  $d$  – количество оцениваемых параметров (Akaike, 1971). Чем меньше, тем лучше.  $AIC \approx 2D_{KL}[f, \widehat{f}_M]$ .

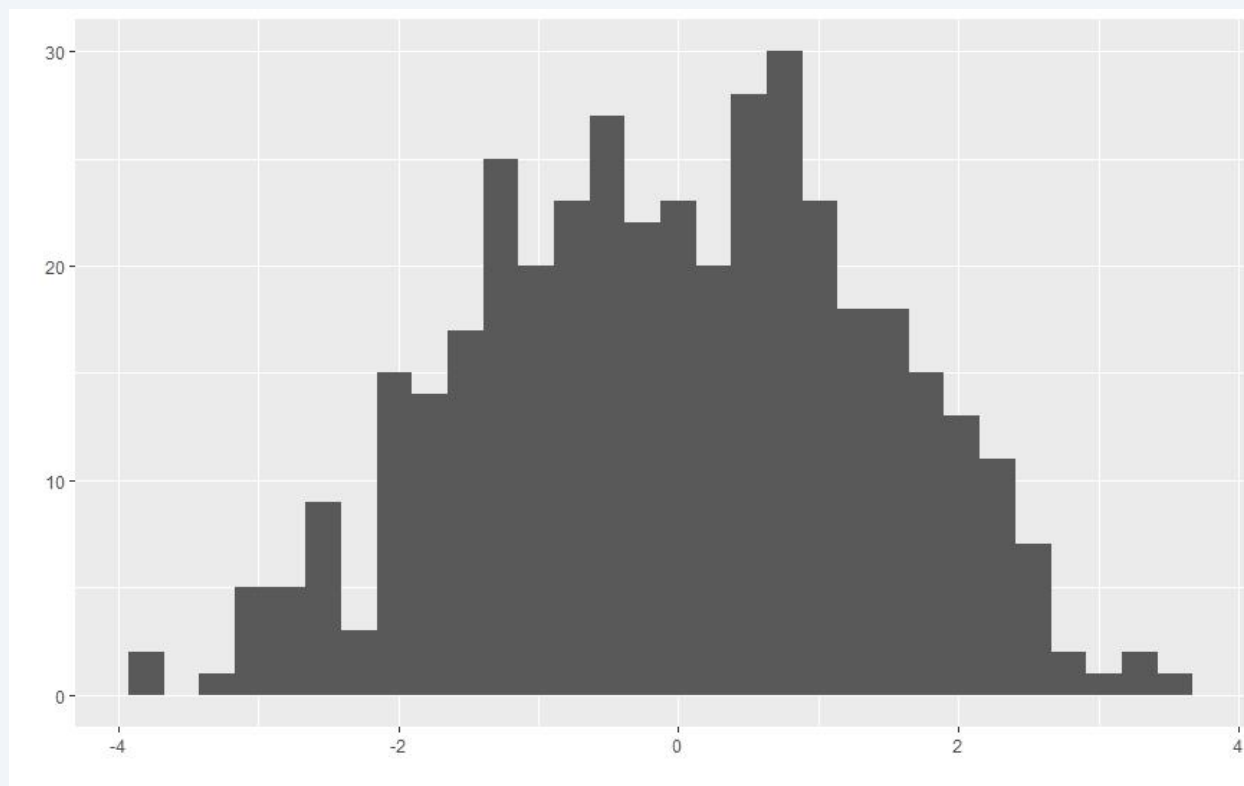
$BIC = 2 \ln \psi(x; \hat{\theta}(x)) - d \ln n$  (Schwartz, 1978). Чем больше, тем лучше.

## Определение границ.

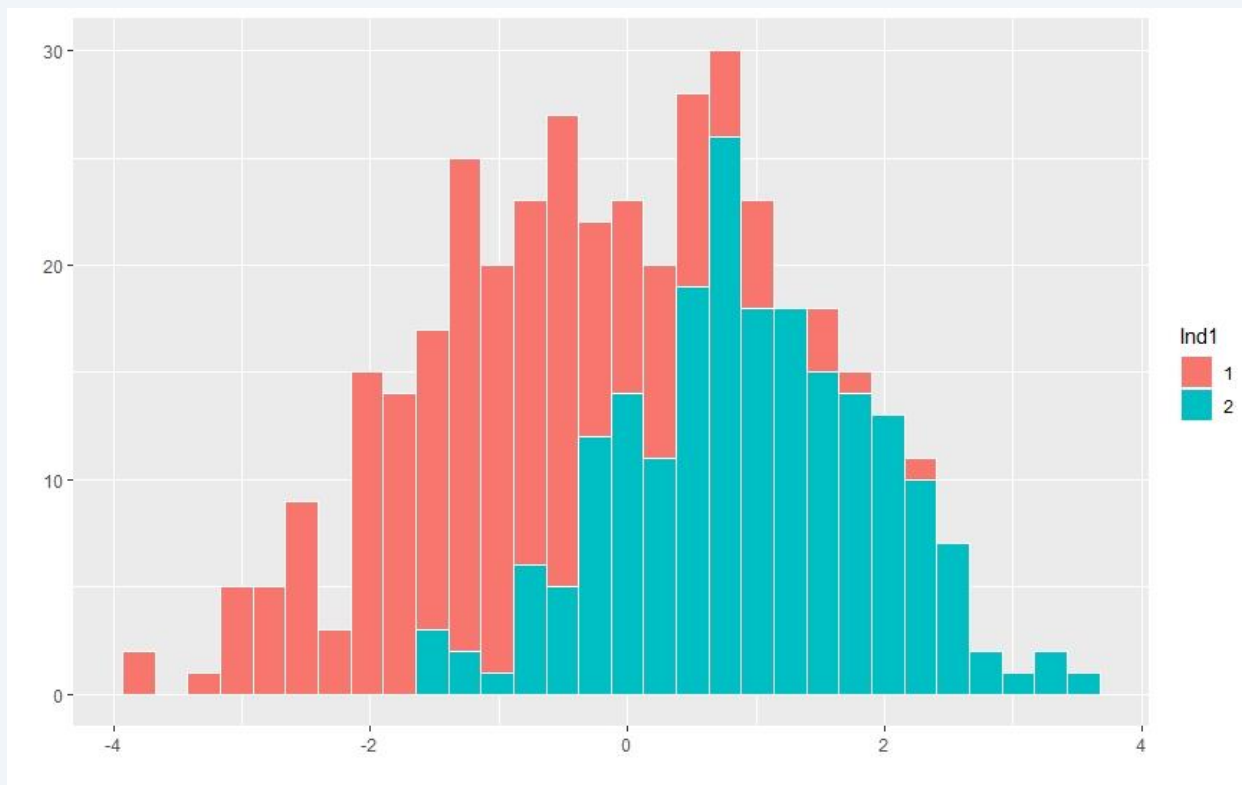


Наблюдение  $x$  отнесем к  $i$ -ой компоненте смеси, если  $p_i f_i(x) > p_j f_j(x)$  для всех  $j \neq i$ .

## Гистограмма смоделированной смеси распределения.



## Гистограмма смеси с выделенными компонентами





## Пример.

Количество уведомлений о смертях женщин старше 80 лет, появляющихся в The Times, в течение трех лет (Everitt B.S. and Hand D.J., 1981).

0	1	2	3	4	5	6	7	8	9
162	267	271	185	111	61	27	8	3	1

Предполагается, что распределение указанных выше смертей описывается распределением Пуассона или смесью пуассоновских распределений.

ОММ и ОМП (одна компонента):  $\lambda^* = 2,16$ ,  $\chi^2 = 26,97$

ОММ (две компоненты):  $\lambda_1^* = 1,1$ ,  $p_1^* = 0,29$ ,  $\lambda_2^* = 2,58$ ,  $\chi^2 = 1,52$

ОМП (две компоненты):  $\lambda_1^* = 1,25$ ,  $p_1^* = 0,36$ ,  $\lambda_2^* = 2,66$ ,  $\chi^2 = 1,19$

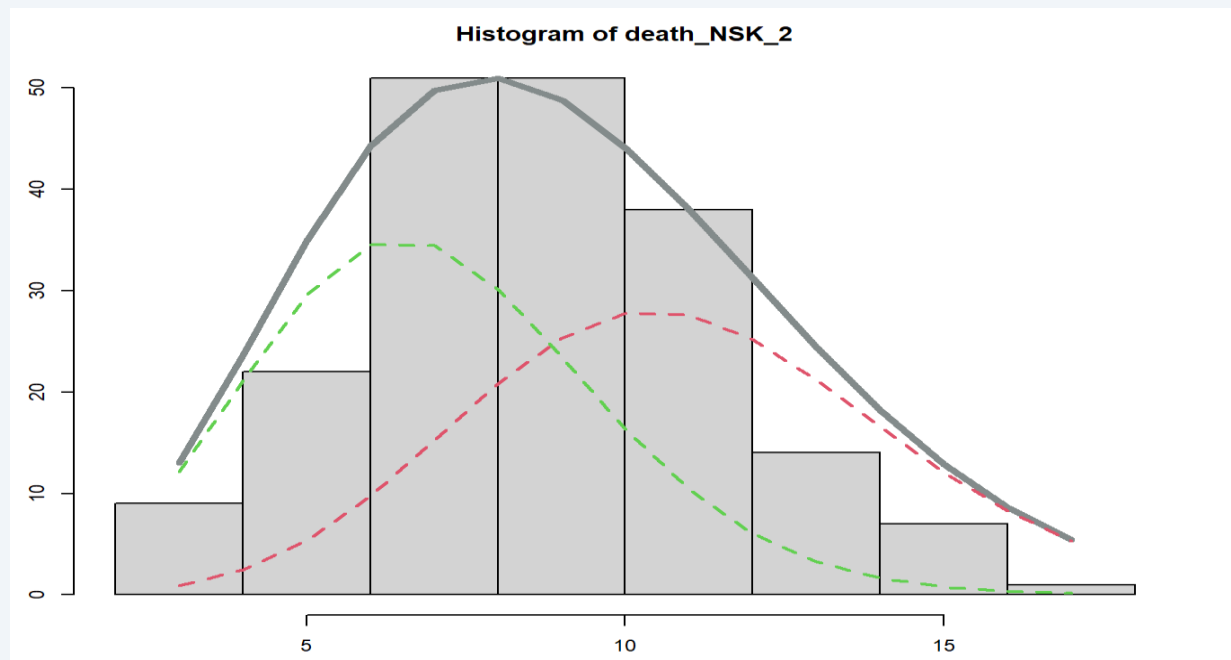
**Грязное статистическое наблюдение.** Рассматривается количество смертей от Covid-19, которые были зафиксированны с лагом две недели в период с 19 октября 2020 по 3 января 2021 и 21 июня 2021 по 13 октября 2021 года. Количество зафиксированный заболеваний в выбранные периоды составило от 150 до 200.

3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
3	6	7	15	28	23	34	17	23	15	13	1	5	2	1

Среднее количество умерших в первый период составило 6.69 (40%)

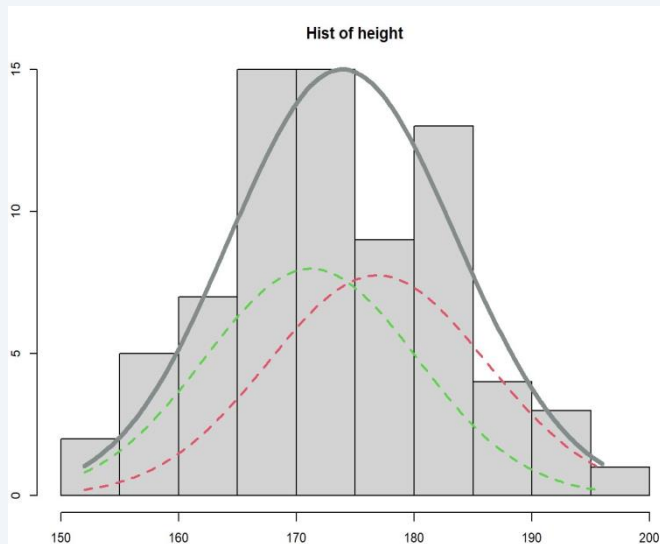
Среднее количество умерших во второй период составило 10.68 (60%)

## Декомпозиция двухкомпонентной смеси пуассоновских распределений

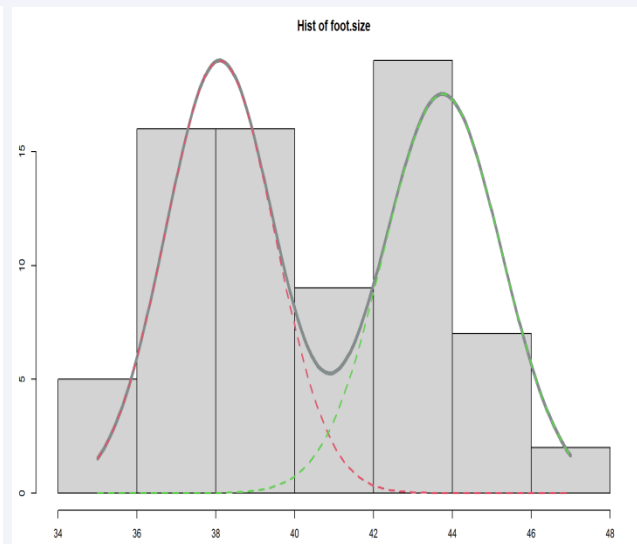


Cluster: 1	Cluster: 2
Proportion = 0.4700124	Proportion = 0.5299876
lambda = 6.993936	lambda = 10.94525

**Статистическое наблюдение.** Были собраны данные о росте и размере ноги у 37 юношей и 37 девушек (студенты 2го курса, 2021 г.).



Cluster: 1	Cluster: 2
Proportion = 0.5004852	Proportion = 0.4995148
Means = 176.8744	Means = 171.2282
S.D. = 9.289872	S.D. = 9.014519



Cluster: 1	Cluster: 2
Proportion = 0.5035358	Proportion = 0.4964642
Means = 38.09885	Means = 43.75193
S.D. = 1.380312	S.D. = 1.49548

## Предпосылки

Для социально-экономических явлений и процессов характерна неоднородность. Это может быть следствием как естественных причин, когда условия формирования показателя настолько многообразны, что в результате развитие различных единиц совокупности может идти по разным законам, так и искусственных причин, когда показатель формируется под воздействием льгот, скидок, преференций, запретов или ограничений, установленных в законодательном и нормативном порядке для управления развитием социально-экономических субъектов и регулирования процессов их дифференциации. И вводя критерии и пороги, субъект управления таким образом искусственно создает неоднородную структуру в попытке задать вектор развития отдельных типов общей совокупности. В таких случаях установление границ – порогов совокупности – носит субъективный характер (не всегда имеет место содержательное обоснование его величины). И, в связи с этим одной из важнейших задач исследования является задача определения объективных правил и критериев нахождения пороговых значений (определение границ типов), чтобы они четко обозначали переходы одного качественного состояния явления в другое.

## Анализ уровня безработицы в субъектах Российской Федерации

Динамика уровня безработицы в Российской Федерации за 2005-2020 гг.



## Оценки параметров смесей нормальных распределений

Год	$p_1$	$a_1$	$\sigma_1$	$p_2$	$a_2$	$\sigma_2$	$p_3$	$a_3$	$\sigma_3$
2005	0,82	7,66	2,09	0,18	23,37	19,20	-	-	-
2010	0,98	8,10	2,76	0,02	46,50	4,53	-	-	-
2011	0,98	7,23	2,44	0,02	42,70	7,64	-	-	-
2012	0,85	5,83	1,76	0,15	22,05	14,34	-	-	-
2013	0,85	5,75	1,76	0,15	20,92	12,68	-	-	-
2014	0,93	5,64	1,96	0,07	20,85	6,96	-	-	-
2015	0,86	6,07	1,91	0,14	20,33	6,93	-	-	-
2016	0,85	6,07	2,01	0,15	17,80	7,15	-	-	-
2017	0,75	5,18	1,26	0,25	11,43	5,28	-	-	-
2018	0,69	4,81	0,98	0,31	8,78	4,91	-	-	-
2019	0,70	4,55	0,85	0,30	8,71	4,97	-	-	-
2020	0,72	5,36	1,10	0,15	8,32	0,92	0,13	15,37	5,96



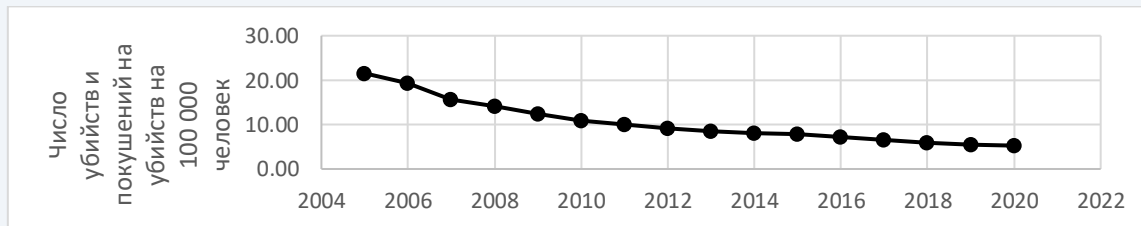
## «Безработные» субъекты Российской Федерации

Год	Порог	Субъект Российской Федерации
<b>2005</b>	13,5	Республика Калмыкия, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Тыва
<b>2010</b>	24	Республика Ингушетия, Чеченская Республика
<b>2011</b>	18,2	Республика Ингушетия, Чеченская Республика
<b>2012</b>	10,9	Республика Калмыкия, Республика Дагестан, Республика Ингушетия, Чеченская Республика, Республика Алтай, Республика Тыва
<b>2013</b>	10,8	Республика Калмыкия, Республика Дагестан, Республика Ингушетия, Чеченская Республика, Республика Алтай, Республика Тыва
<b>2014</b>	11,7	Республика Ингушетия, Чеченская Республика
<b>2015</b>	11,4	Республика Ингушетия, Карачаево-Черкесская Республика, Чеченская Республика, Республика Тыва
<b>2016</b>	11,4	Республика Ингушетия, Карачаево-Черкесская Республика, Чеченская Республика, Республика Алтай, Республика Тыва
<b>2017</b>	8,12	Республика Карелия, Республика Адыгея, Республика Калмыкия, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания, Чеченская Республика, Курганская область, Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Иркутская область, Еврейская автономная область

Год	Порог	«Промежуточные» субъекты РФ	Порог	«Безработные» субъекты РФ
2018	-	?	6,75	Республика Карелия, Республика Коми, Республика Адыгея, Республика Калмыкия, Астраханская область, Республика Дагестан, <b>Республика Ингушетия</b> , Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия – Алания, Чеченская Республика, Курганская область, Республика Алтай, Республика Тыва, Иркутская область, Республика Бурятия, Забайкальский край, Еврейская автономная область
2019	-	?	6,43	Республика Карелия, Республика Коми, Республика Адыгея, Республика Калмыкия, Астраханская область, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия – Алания, Чеченская Республика, Курганская область, Республика Алтай, Республика Тыва, Иркутская область, Республика Бурятия, Республика Саха (Якутия), Забайкальский край
2020	7,16	Республика Карелия, Республика Коми, Мурманская область, Республика Адыгея, Республика Калмыкия, Астраханская область, Волгоградская область, Курганская область, Республика Хакасия, Иркутская область, Омская область, Томская область, Забайкальский край	10,21	Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия – Алания, Чеченская Республика, Республика Алтай, Республика Тыва, Республика Бурятия

## Анализ криминогенной ситуации в субъектах Российской Федерации

Динамика числа убийств и покушений на убийство на 100 000 человек в Российской Федерации за 2005-2020 гг.



Оценки параметров двухкомпонентных смесей нормальных распределений

Год	$p$	$a_1$	$\sigma_1$	$a_2$	$\sigma_2$
2010	0,79	11,27	3,75	30,27	13,41
2011	0,77	10,06	3,33	25,72	7,95
2012	0,78	9,32	3,26	26,12	9,84
2013	0,69	7,55	1,90	18,14	6,41
2014	0,79	8,26	3,36	23,54	8,94
2015	0,79	8,17	3,25	23,37	7,88
2016	0,77	6,68	2,41	18,81	6,75
2017	0,74	6,09	2,06	16,00	5,26

## «Криминальные» субъекты Российской Федерации

Год	Порог	Субъект Российской Федерации
2010	20,23	Ненецкий автономный округ, Республика Алтай, <i>Республика Бурятия, Республика Тыва, Забайкальский край</i> , Иркутская область
2011	17,67	Республика Алтай, <i>Республика Бурятия, Республика Тыва, Забайкальский край</i> , Иркутская область, Амурская область, Магаданская область, Еврейская автономная область, Чукотский автономный округ
2012	17,03	<i>Республика Бурятия, Республика Тыва, Забайкальский край</i> , Иркутская область, Амурская область, Сахалинская область, Еврейская автономная область, Чукотский автономный округ
2013	11,79	Республика Коми, Ненецкий автономный округ, Новгородская область, Карачаево-Черкесская Республика, Пермский край, Курганская область, Курганская область, Ямало-Ненецкая автономная область, Республика Алтай, <i>Республика Бурятия, Республика Тыва</i> , Республика Хакасия, <i>Забайкальский край</i> , Иркутская область, Кемеровская область, Республика Саха (Якутия), Приморский край, Хабаровский край, Амурская область, Магаданская область, Сахалинская область, Еврейская автономная область, Чукотский автономный округ
2014	16	Республика Алтай, <i>Республика Бурятия, Республика Тыва, Забайкальский край</i> , Иркутская область, Приморский край, Магаданская область, Чукотский автономный округ
2015	15,74	Ненецкий автономный округ, Республика Алтай, <i>Республика Бурятия, Республика Тыва, Забайкальский край</i> , Магаданская область, Сахалинская область
2016	12,28	Республика Коми, Архангельская область, Ненецкий автономный округ, Республика Алтай, <i>Республика Бурятия, Республика Тыва</i> , Республика Хакасия, <i>Забайкальский край</i> , Иркутская область, Республика Саха (Якутия), Хабаровский край, Амурская область, Магаданская область, Еврейская автономная область, Чукотский автономный округ
2017	10,68	Архангельская область, Ненецкий автономный округ, Курганская область, Республика Алтай, <i>Республика Бурятия, Республика Тыва</i> , Республика Хакасия, <i>Забайкальский край</i> , Иркутская область, Кемеровская область, Республика Саха (Якутия), Хабаровский край, Амурская область, Магаданская область, Сахалинская область, Еврейская автономная область, Чукотский автономный округ

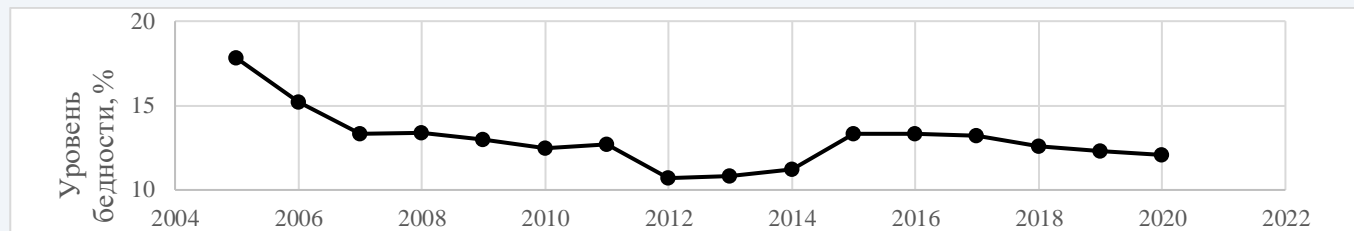
Год	Порог	«Промежуточные» субъекты РФ	Порог	«Криминальные» субъекты РФ
2018	9,22	Республика Коми, Архангельская область, Курганская область, Республика Хакасия, Красноярский край, Иркутская область, Республика Саха (Якутия), Камчатский край, Хабаровский край, Сахалинская область	12,40	Республика Алтай, Республика Тыва, Республика Бурятия, Забайкальский край, Амурская область, Магаданская область, Еврейская автономная область, Чукотский автономный округ
2019	5,87	31 субъект	12,04	Республика Тыва, Республика Бурятия, Забайкальский край, Амурская область, Магаданская область, Еврейская автономная область, Чукотский автономный округ
2020	6,20	21 субъект	14,96	Республика Тыва, Республика Бурятия, Забайкальский край, Чукотский автономный округ

### Оценки параметров трехкомпонентных смесей нормальных распределений

Год	$p_1$	$a_1$	$\sigma_1$	$p_2$	$a_2$	$\sigma_2$	$p_3$	$a_3$	$\sigma_3$
2018	0,77	5,18	1,69	0,10	10,91	0,73	0,13	15,66	5,05
2019	0,40	4,24	1,07	0,48	6,46	2,31	0,12	17,66	7,84
2020	0,62	4,36	1,28	0,30	8,65	2,95	0,07	17,09	7,18

## Анализ уровня бедности в субъектах Российской Федерации

Динамика уровня бедности в Российской Федерации за 2005-2020 гг.



Оценки параметров двухкомпонентных смесей нормальных распределений

Год	$p$	$a_1$	$\sigma_1$	$a_2$	$\sigma_2$
2005	0,92	20,43	5,74	46,27	9,78
2010	0,97	14,63	3,82	32,65	4,31
2011	0,98	14,66	3,89	33,20	3,68
2012	0,98	12,75	3,51	29,30	1,98
2013	0,98	12,93	3,49	34,40	1,41
2015	0,96	14,42	3,99	34,53	3,40
2016	0,86	13,91	3,40	28,06	6,25
2017	0,87	13,74	3,48	27,76	5,89
2018	0,81	13,00	3,26	20,86	6,18
2019	0,79	12,86	3,15	20,68	6,06
2020	0,80	12,53	3,17	20,52	5,60



## «Бедные» субъекты Российской Федерации

Год	Порог	Субъект Российской Федерации
<b>2005</b>	35,63	Ивановская область, <i>Республика Калмыкия</i> , Республика Ингушетия, Республика Марий Эл, Республика Алтай, <i>Республика Тыва</i>
<b>2010</b>	26,45	<i>Республика Калмыкия</i> , <i>Республика Тыва</i>
<b>2011</b>	27,02	<i>Республика Калмыкия</i> , <i>Республика Тыва</i>
<b>2012</b>	24,72	<i>Республика Калмыкия</i> , <i>Республика Тыва</i>
<b>2013</b>	28,88	<i>Республика Калмыкия</i> , <i>Республика Тыва</i>
<b>2015</b>	27,44	<i>Республика Калмыкия</i> , Республика Ингушетия, <i>Республика Тыва</i>
<b>2016</b>	22,11	<i>Республика Калмыкия</i> , Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Марий Эл, Республика Алтай, <i>Республика Тыва</i> , Еврейская автономная область
<b>2017</b>	22,07	<i>Республика Калмыкия</i> , Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Марий Эл, Республика Алтай, <i>Республика Тыва</i> , Еврейская автономная область
<b>2018</b>	19,54	Республика Калмыкия, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Чеченская Республика, Республика Марий Эл, Республика Алтай, Республика Тыва, Забайкальский край, Еврейская автономная область
<b>2019</b>	18,93	Республика Калмыкия, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Чеченская Республика, Республика Марий Эл, Курганская область, Республика Алтай, Республика Тыва, Республика Бурятия, Забайкальский край, Еврейская автономная область
<b>2020</b>	18,78	Республика Калмыкия, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Чеченская Республика, Республика Марий Эл, Курганская область, Республика Алтай, Республика Тыва, Республика Бурятия, Забайкальский край, Еврейская автономная область



#### - области

- |                   |                    |                   |
|-------------------|--------------------|-------------------|
| 1 - Астраханская  | 9 - Ленинградская  | 17 - Рязанская    |
| 2 - Владимирская  | 10 - Липецкая      | 18 - Самарская    |
| 3 - Волгоградская | 11 - Московская    | 19 - Свердловская |
| 4 - Воронежская   | 12 - Нижегородская | 20 - Тамбовская   |
| 5 - Ивановская    | 13 - Новгородская  | 21 - Тульская     |
| 6 - Калужская     | 14 - Орловская     | 22 - Ульяновская  |
| 7 - Кемеровская   | 15 - Пензенская    | 23 - Челябинская  |
| 8 - Костромская   | 16 - Ростовская    | 24 - Ярославская  |

#### - республики

- |                          |                              |
|--------------------------|------------------------------|
| 1 - Адыгея               | 8 - Мордовия                 |
| 2 - Башкортостан         | 9 - Северная Осетия - Алания |
| 3 - Ингушетия            | 10 - Татарстан               |
| 4 - Кабардино-Балкарская | 11 - Удмуртская              |
| 5 - Калмыкия             | 12 - Хакасия                 |
| 6 - Карачаево-Черкесская | 13 - Чеченская               |
| 7 - Марий Эл             | 14 - Чувашская               |

#### - автономные округа

- |                                  |
|----------------------------------|
| - края                           |
| 1 - Ставропольский               |
| - города федерального назначения |
| 1 - Москва                       |
| - автономная область             |

■ Высокий уровень числа убийств и бедности  
 ■ Высокий уровень безработицы  
 ■ Высокий уровень бедности  
 ■ Высокий уровень числа убийств

The map displays the following regional distributions:

- High level of crime and poverty (Purple):** Chuvashia.
- High level of unemployment (Yellow):** Dagestan.
- High level of poverty (Red):** Chechnya.
- High level of crime (Blue):** Chukotka, Khanty-Mansi Autonomous Okrug, Magadan Oblast, and a large cluster in the Far East including Kamchatka, Khabarovsk Krai, and the Jewish Autonomous Oblast.

Высокий уровень бедности  
 Высокий уровень числа убийств, бедности и безработицы  
 Высокий уровень безработицы  
 Высокий уровень бедности и безработицы  
 Высокий уровень числа убийств

## Типология субъектов РФ по уровню бедности, уровню безработицы и уровню насилия в 2017 г.



## Типология субъектов РФ по уровню бедности, уровню безработицы и уровню насилия в 2020 г.

<b>«Безработные», «бедные» и «криминальные» регионы РФ</b>	Республика Тыва, Республика Бурятия
<b>«Безработные», «бедные», но не «криминальные» регионы РФ</b>	Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Чеченская Республика, Республика Алтай
<b>«Бедные», «криминальные», но не «безработные» регионы РФ</b>	Забайкальский край
<b>«Безработные», но не «бедные» и не «криминальные» регионы РФ</b>	Республика Дагестан, Республика Северная Осетия – Алания
<b>«Бедные», но не «безработные» и не «криминальные» регионы РФ</b>	Республика Калмыкия, Республика Марий Эл, Курганская область, Еврейская автономная область
<b>«Криминальные», но не «бедные» и не «безработные» регионы РФ</b>	Чукотский автономный округ
<b>«Промежуточные» регионы РФ</b>	Республика Карелия, Республика Коми, Мурманская область, Республика Адыгея, Астраханская область, Волгоградская область, Республика Хакасия, Иркутская область, Омская область, Томская область



■ Высокий уровень всех трех показателей  
 ■ Высокий уровень двух показателей  
 ■ Высокий уровень одного показателя  
 ■ Попадание показателя в "промежуточный" тип

## **Вывод**

Несмотря на общую тенденцию к снижению значения всех показателей по Российской Федерации и в отдельных регионах, число регионов, относящихся ко второй или третьей компоненте в смеси, увеличивается. Также увеличивается число регионов, которые характеризуются высокими значениями двух показателей из трех, а также число регионов, которые попадают в типы с высокими значениями всех трех показателей.

## **Заключение.**

Типологическая группировка реальных социально-экономических совокупностей может быть использована при разработке и корректировке органами государственной власти документов стратегического планирования, региональных и федеральных программ экономического развития, и принятия управленческих решений, направленных на регулирование и контроль дифференциации территорий.

## **Литература.**

Айвазян С. А., Бухштабер В.М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. – Москва, 1989.

Королёв В. Ю. ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор. – М.: ИПИ РАН, 2007 – 94 с.

Akaike H. Information theory and an extension of the maximum likelihood principle. // Second International Symposium on Information Theory / in: B. N. Petrov and F. Csake (eds.).– Akademiai Kiado, Budapest, 1973. – pp. 267-281.

Celeux G. and Diebolt J. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem // Computation Statistics Quarterly, 1985. – vol. 2, No 1. – pp. 73-82.

Dempster A., Laird N. and Rubin D. Maximum likelihood estimation from incomplete data. – Journal of the Royal Statistical Society, Series B, 1977, vol. 39, p. 1-38.

Everitt B.S. and Hand D.J. Finite Mixture Distributions. – London: Chapman & Hall, 1981.

Schwartz G. Estimating the dimension of a model. – The Anaals of Statistics, 1978, vol. 6, p. 461-464.

Pearson K. Contributions to the Mathematical Theory of Evolution // Philosophical Transactions of the Royal Society of London. – 1894. – Vol. 185. –pp. 71-110.