Возможности применения метода АНС на микро-данных ВПН 2010 года

Тындик Алла

Семинар НУГ

Изучение рождаемости, формирования, развития и распада семей на данных выборочных обследований

07.02.2014

Предпосылки анализа

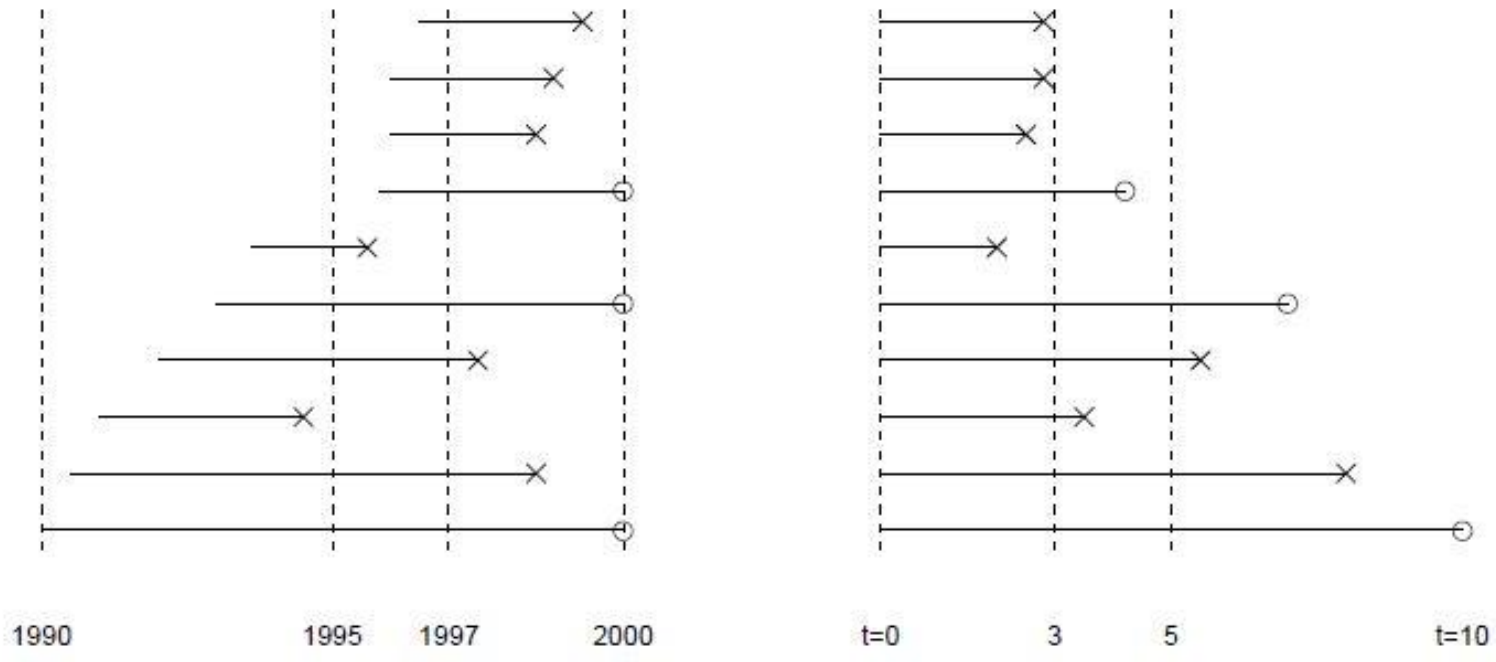
- Откладывание рождения первенца
- Рост доли окончательно бездетных

+

- Новые возможности анализа старого источника данных
- Использование методов анализа наступления событий

Анализ наступления событий

Цензурирование:



The survival function gives the **proportion of people NOT having the event** (e.g. surviving) until time t.

- АНС – изучение зависимости риска наступления (или ненаступления) события от продолжительности пребывания объекта в группе риска, а также от ряда других характеристик объекта и внешних факторов, влияющих на этот риск.
- В пакете SPSS реализованы четыре инструмента для проведения анализа наступления событий: табличный метод (Life Tables), метод Каплана–Мейера (Kaplan–Meier Estimator), регрессия Кокса (Cox Regression) и регрессия Кокса с зависящими от времени объясняющими переменными (Cox Regression with Time Dependent Covariates).
- Все независимые переменные должны быть измерены на момент наступления события (!)
- Статья Бурдяк А.Я. «Применение метода анализа наступления события с помощью пакета SPSS»

Вид переменных

- Событие: 0 – 1
- Время: с точки отсчета до наступления события или до момента цензурирования, заданная единица отсчета
- Переменные-факторы, категориальные

Данные ВПН: первоначальный вид

std.gmcrosstata.ru/webapi/jsf/tableView/customiseTable.xhtml

200 лучших нон-ф... | Академическое пи... | Практики коммем...

SuperWEB2™

Таблица | Диаграмма | Карта*

Настроить таблицу | Мои пользовательские данные | Мои таблицы | Загрузить таблицу Exce

Закрыть

Выбрать тему

Всероссийские переписи населения 2002 2010 годов

Выберите элементы ниже, затем постройте таблицу. Помощь

Добавить к строке | Добавить к столбцу | Удалить из таблицы

Свернуть все | Снять все отметки | 0 элементов выбрано.

- 1995
- 1996
- 1997
- 1998
- 1999
- 2000
- *2001*
- *2002*
- *2003*
- *2004*
- *2005*
- *2006*
- *2007*
- *2008*
- *2009*
- *2010*
- Тип населения [2]
- Возрастная группа [4]
- Число источников средств к существованию [13]
- Признак экономической активности [4]
- Возраст, в котором женщина родила первого ребенка [49]
- Количество языков [10]
- Основной источник дохода [14]

Извлечь данные | Авто | Процент: Отсутствует

Всероссийские переписи населения 2002 2010 годов

Л3 Год рождения по Год переписи и Л13.2 Год рождения первого ребенка

Подсчет: Число лиц

Для получения дополнительных сведений см. Конфиденциальность данных.

Подсчет ячеек таблицы, включая итоги: **242 (22 столбцы x 11 строки).**

Год переписи	2010									
Л13.2 Год рождения первого ребенка	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001
Л3 Год рождения										
1985	58 855	80 989	82 544	83 936	84 256	78 665	57 758	34 927	15 958	5 937
1984	53 825	76 064	79 524	80 360	84 745	84 767	79 892	58 348	33 942	15 555
1983	46 676	69 506	75 377	79 538	84 247	89 439	91 764	84 992	60 937	35 295
1982	35 499	54 341	62 117	68 083	74 860	81 146	88 789	90 173	82 028	57 608
1981	27 485	42 488	49 428	56 263	64 213	72 042	79 869	87 184	87 952	77 995
1980	22 197	34 404	41 116	47 670	56 713	69 105	76 997	85 276	90 831	91 054
1979	17 034	26 386	31 247	36 292	43 875	53 711	64 069	72 708	79 574	83 639
1978	13 416	20 877	24 914	28 712	34 929	42 847	52 954	62 903	71 302	75 777
1977	10 203	16 057	19 380	22 886	26 987	33 298	41 294	50 318	58 952	65 357
1976	7 956	12 734	14 996	17 857	21 538	25 775	32 346	39 602	48 027	55 851

Вы можете настроить таблицу (Семейная ячейка, Домохозяйство, Помещение, Тип населения (город/село), ТЕРСОН-АТЕ, ТЕРСОН-МО, ТЕРСОН-
выбрав Настроить таблицу .

INFO: К значениям в таблице был применен механизм защиты конфиденциальной информации.

Преобразование данных

	A	B	C	D
1	birthR	birthCh	cases100	id1
2	1950	1965	8	1086
3	1950	1966	22	1087
4	1950	1967	81	1088
5	1950	1968	256	1089
6	1950	1969	568	1090
7	1950	1970	1121	1091
8	1950	1971	1288	1092
9	1950	1972	1320	1093
10	1950	1973	1078	1094
11	1950	1974	857	1095
12	1950	1975	682	1096
13	1950	1976	468	1097
14	1950	1977	335	1098
15	1950	1978	272	1099
16	1950	1979	211	1100
17	1950	1980	198	1101
18	1950	1981	132	1102
19	1950	1982	111	1103
20	1950	1983	95	1104
21	1950	1984	75	1105
22	1950	1985	68	1106
23	1950	1986	56	1107
24	1950	1987	46	1108
25	1950	1988	34	1109
26	1950	1989	27	1110
27	1950	1990	20	1111
28	1950	1991	13	1112
29	1950	1992	8	1113
30	1950	1993	5	1114
31	1950	1994	3	1115
32	1950	1995	2	1116
33	1951	1966	4	1117
34	1951	1967	19	1118
35	1951	1968	83	1119

	A	B	C	D	E
1	birthR	birthCh	cases100	id1	real_id
2	1950	1965	8	1086	10861
3	1950	1965	8	1086	10862
4	1950	1965	8	1086	10863
5	1950	1965	8	1086	10864
6	1950	1965	8	1086	10865
7	1950	1965	8	1086	10866
8	1950	1965	8	1086	10867
9	1950	1965	8	1086	10868
10	1950	1966	22	1087	10871
11	1950	1966	22	1087	10872
12	1950	1966	22	1087	10873
13	1950	1966	22	1087	10874
14	1950	1966	22	1087	10875
15	1950	1966	22	1087	10876
16	1950	1966	22	1087	10877
17	1950	1966	22	1087	10878
18	1950	1966	22	1087	10879
19	1950	1966	22	1087	108710
20	1950	1966	22	1087	108711
21	1950	1966	22	1087	108712
22	1950	1966	22	1087	108713
23	1950	1966	22	1087	108714
24	1950	1966	22	1087	108715
25	1950	1966	22	1087	108716
26	1950	1966	22	1087	108717
27	1950	1966	22	1087	108718
28	1950	1966	22	1087	108719
29	1950	1966	22	1087	108720

+ аналогично по нерожавшим

АНС на данных переписи

Массив:

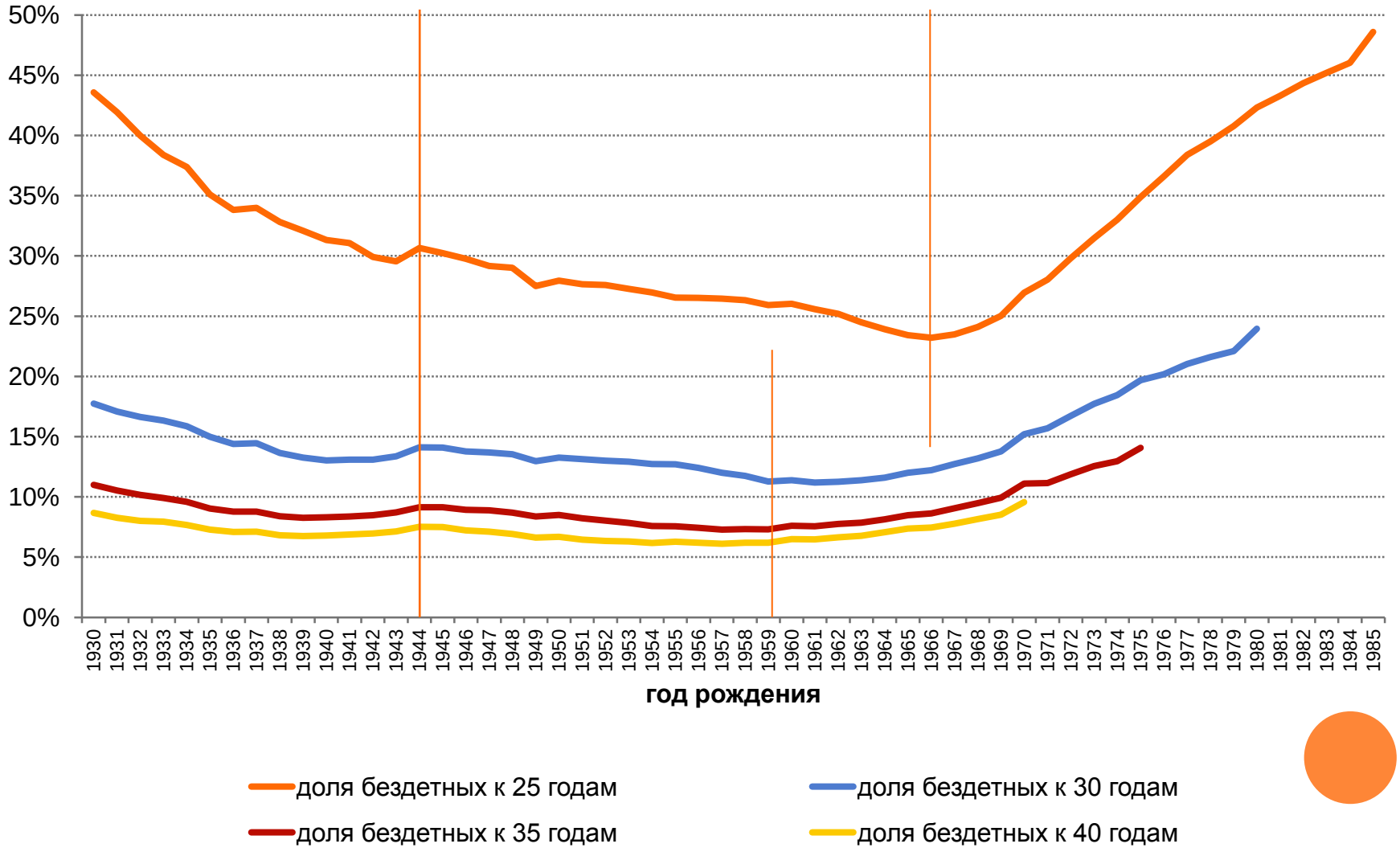
- Ограничения: масштаб 1 кейс = 100 женщин
- Ограничения: используются только года
- Женщины 1926-1995 гг. рождения («с запасом»)
- Объем: **611227 кейсов** (61,1 млн женщин)
- Доля событий: 78,3% родили первого ребенка

АНС:

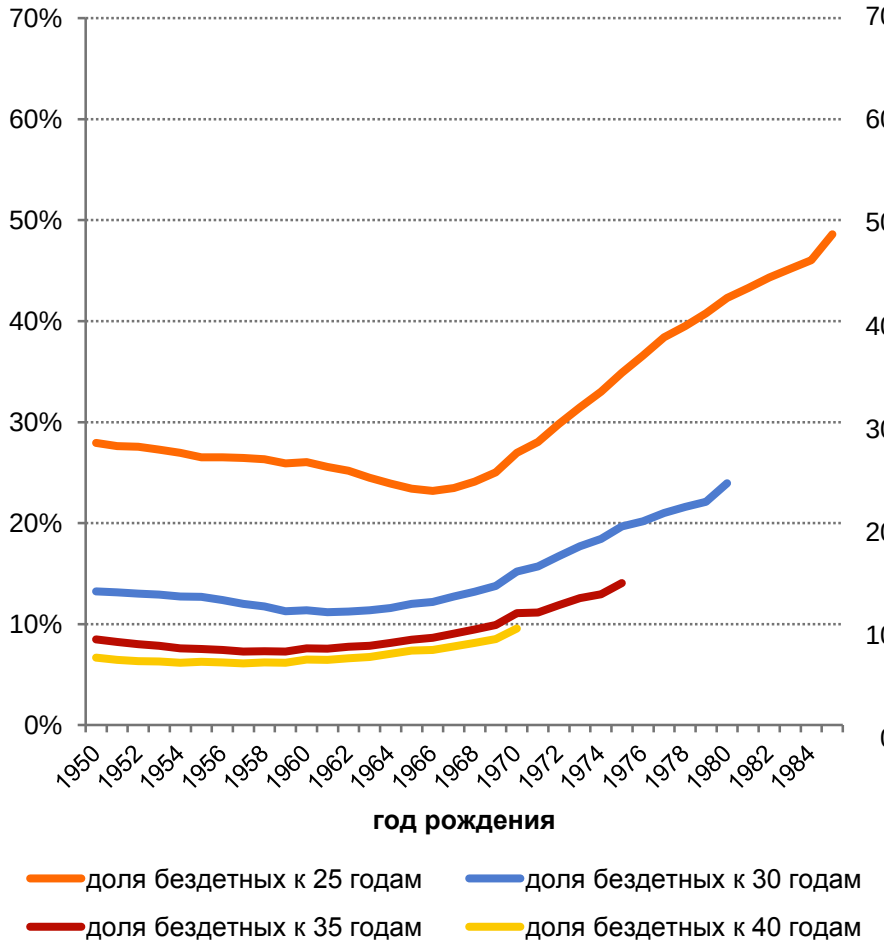
- Старт времени: 15 лет
- Отсчет: погодично
- Финиш: год рождения первого ребенка ИЛИ

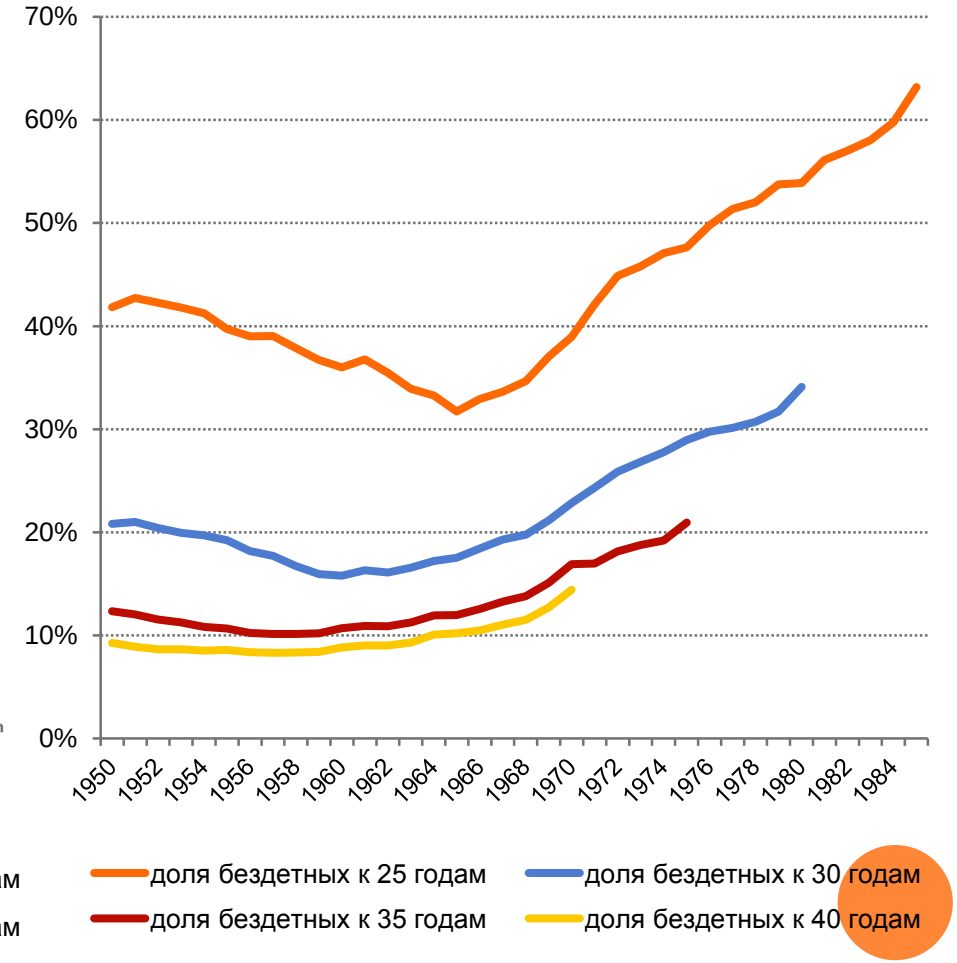
45 лет ИЛИ возраст на момент переписи

Рост доли бездетных

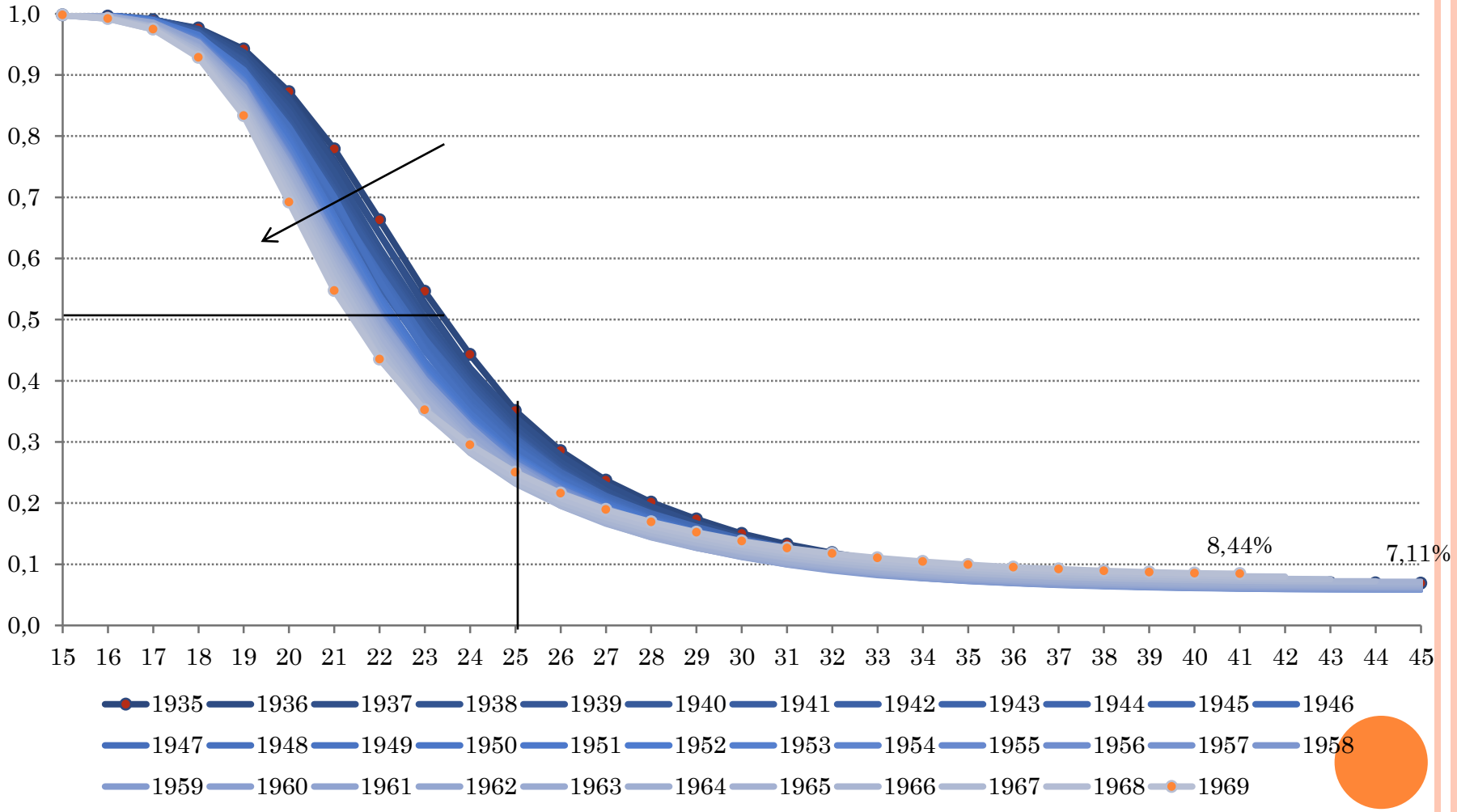


Россия в целом vs г. Москва



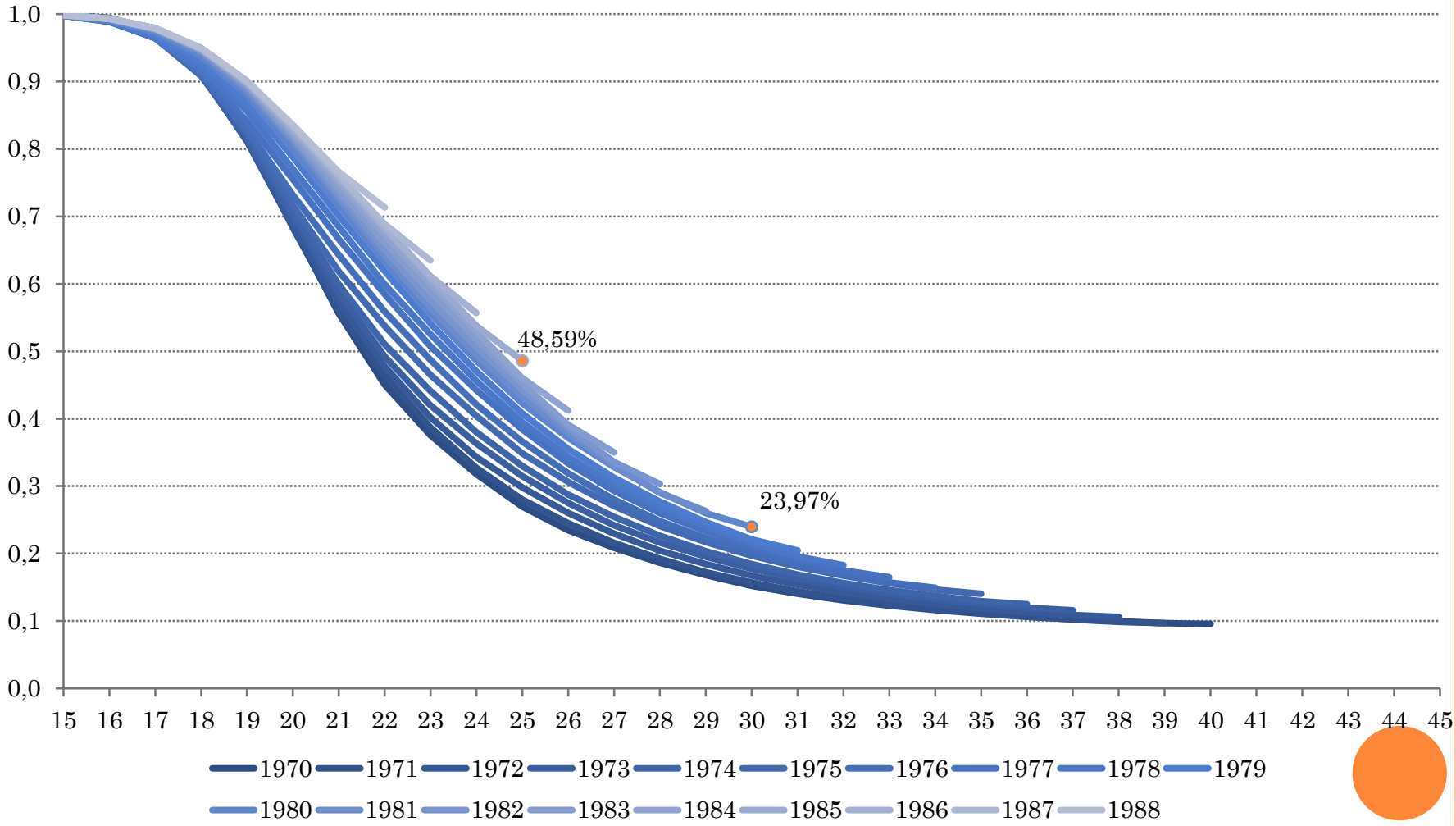


АНС методом Каплана-Майера: когорты с 1935 по 1969 гг. рождения

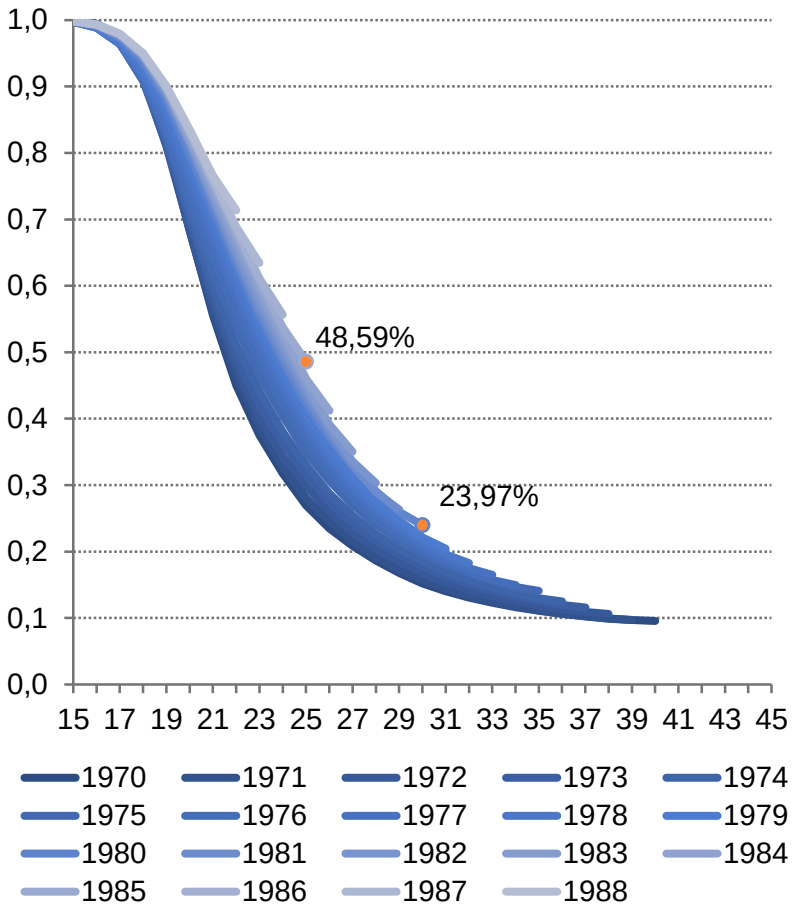


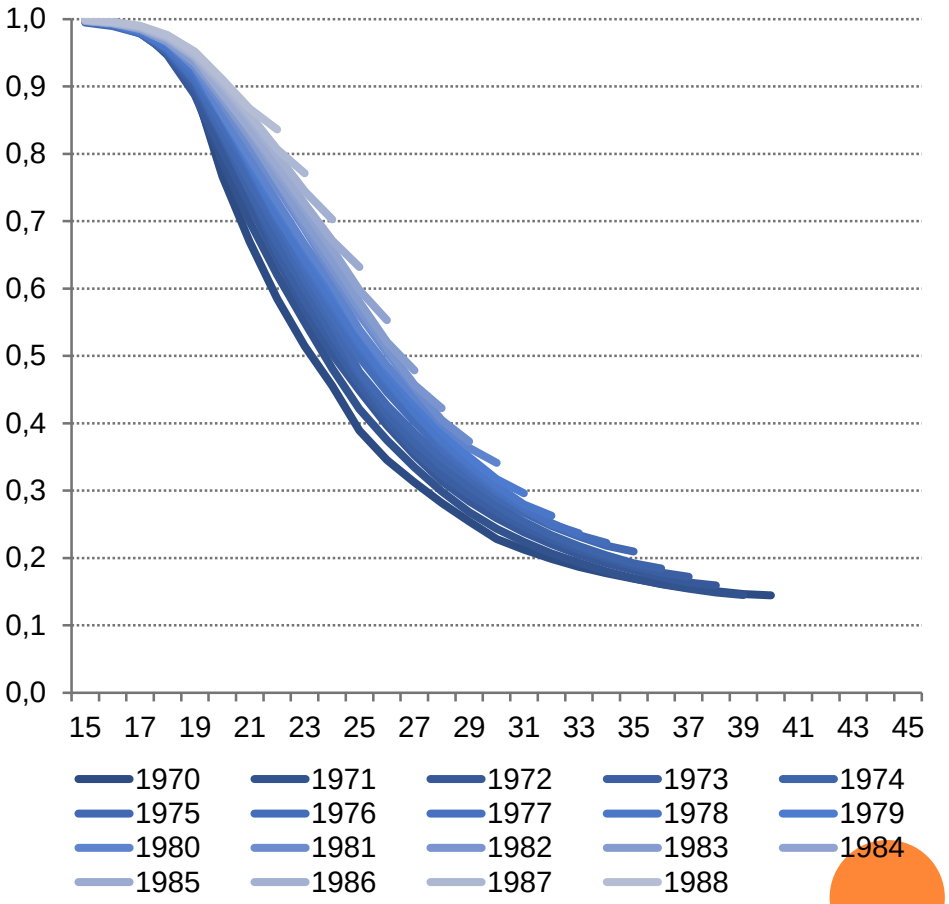
1969 + 22,6 = 1991

АНС методом Каплана-Майера: когорты с 1970 по 1988 гг. рождения



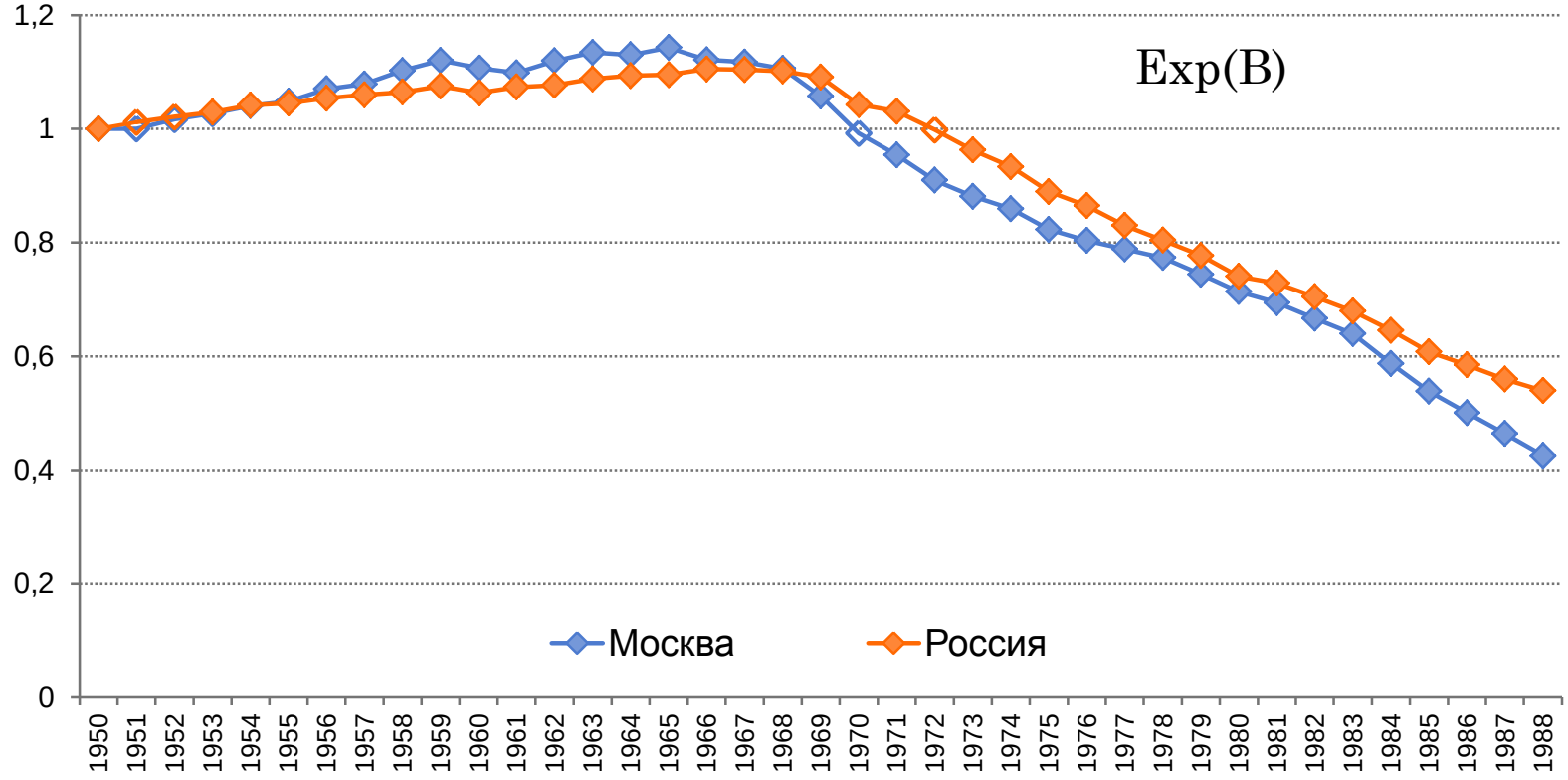
Россия в целом vs г. Москва





Регрессия Кокса
(прогнозирование риска наступления события и оценка влияния года рождения на этот риск)

При увеличении г.р. на единицу риск наступления события возрастает в exp(β) раз



1950 г.р.: рождение первого ребенка состоялось в 93,6% случаев (9459 кейсов)
1988 г.р.: рождение первого ребенка состоялось в 28,7% случаев (3136 кейсов)

Дальнейшие возможности анализа

- Разделение женщин по группам (трудоемко в силу формата исходных данных). Можно использовать только ретроспективные факторы:
  - национальность
  - регион рождения
  - разбивка на территории любого уровня
  - возможно включение миграции в качестве зависимой от времени переменной – есть год последнего переезда (например, влияние факта и возраста переезда в Москву на календарь первых рождений)
- Коррекция данных?..
- Автоматизация