

Богатые реже плачут

(или почему счастье в деньгах)

Panic! At The Kernel

Сафонова Арина Эйбатов Эмиль

Солопов Евгений Пильноватых Глеб

Депрессия – очень распространённое заболевание

World Health Organization^[1]

*"Globally, it is estimated that **4.4%** of the global population suffer from depressive disorder"*

*"Depressive disorders led to a global total of over 50 million Years Lived with Disability (YLD) in 2015. Globally, depressive disorders are ranked as the single largest contributor to non-fatal health loss (**7.5%** of all YLD)..."*

Если

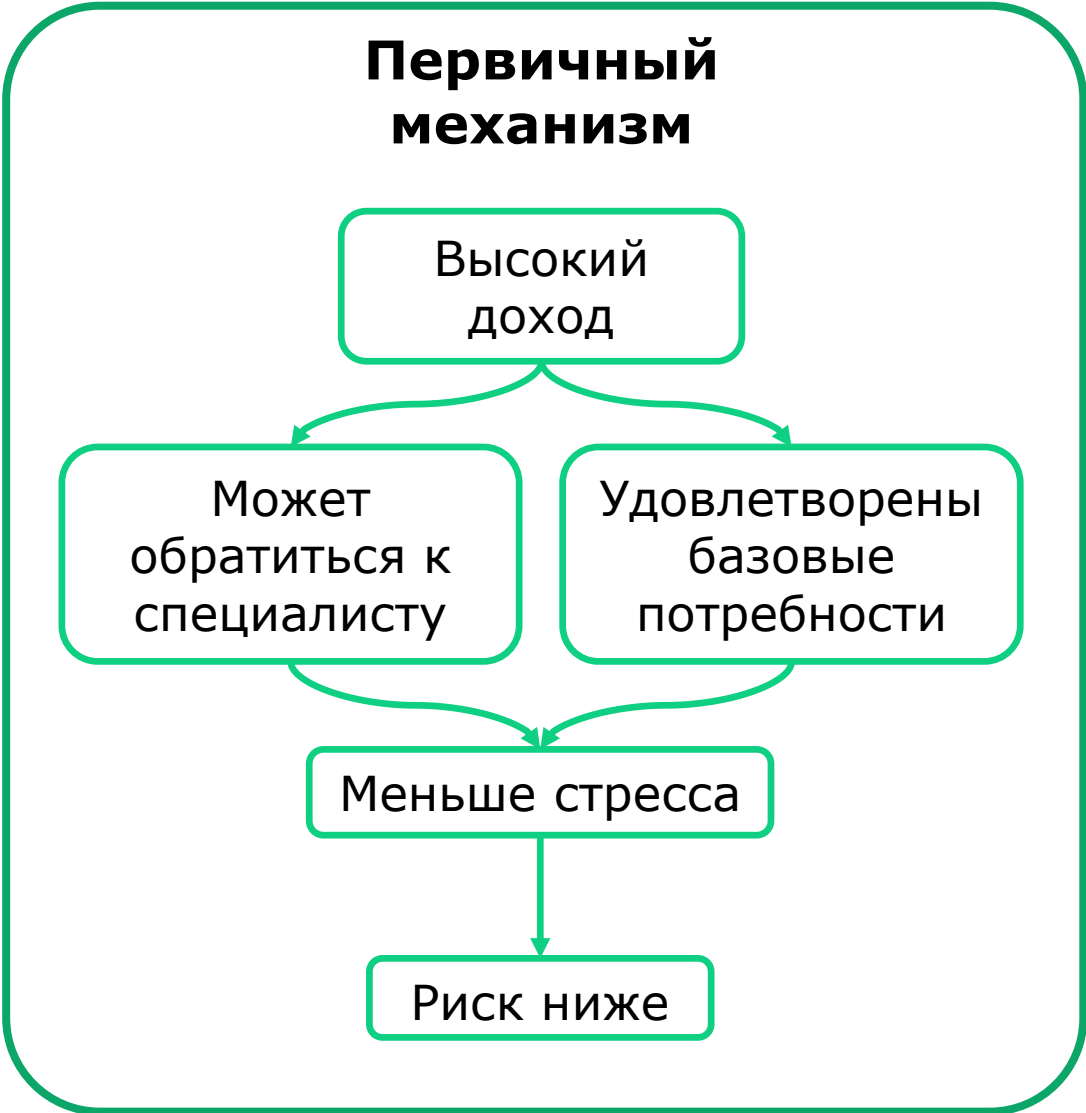
Высокий доход позволяет
смягчить факторы,
приводящие к
депрессивным симптомам



Возможна адаптация
социально-экономических
программ поддержки и
предотвращения

[1] World Health Organization et al. (2017). *Depression and other common mental disorders: global health estimates*. Tech. rep. World Health Organization.

Как доход индивида влияет на вероятность наличия депрессивных симптомов?



От механизмов к гипотезам

Гипотеза 1

В преклонном возрасте у
более богатых людей
реже бывают
депрессивные симптомы*

Гипотеза 2

Влияние дохода на
наличие депрессивных
симптомов не изменилось
за годы опроса

*Ряд исследований подтверждает наличие
закономерности на иных генеральных совокупностях.
Однако поведение зависимости заметно отличается^[2]

[2] [Sareen et al. 2011] [Shields-Zeeman and Smit 2022] [Ridley et al. 2020] [Shields-Zeeman, Collin, et al. 2021] [Gresenz et al. 2001] [Weinberger et al. 2018]

Был задействован изначальный датасет RAND RHS

Дизайн опроса

- Смешанный опрос (Интервьюирование и Анкетирование)
- Панельный вид
- 1991-2012 год
- Открыт доступ к шаблонам анкет

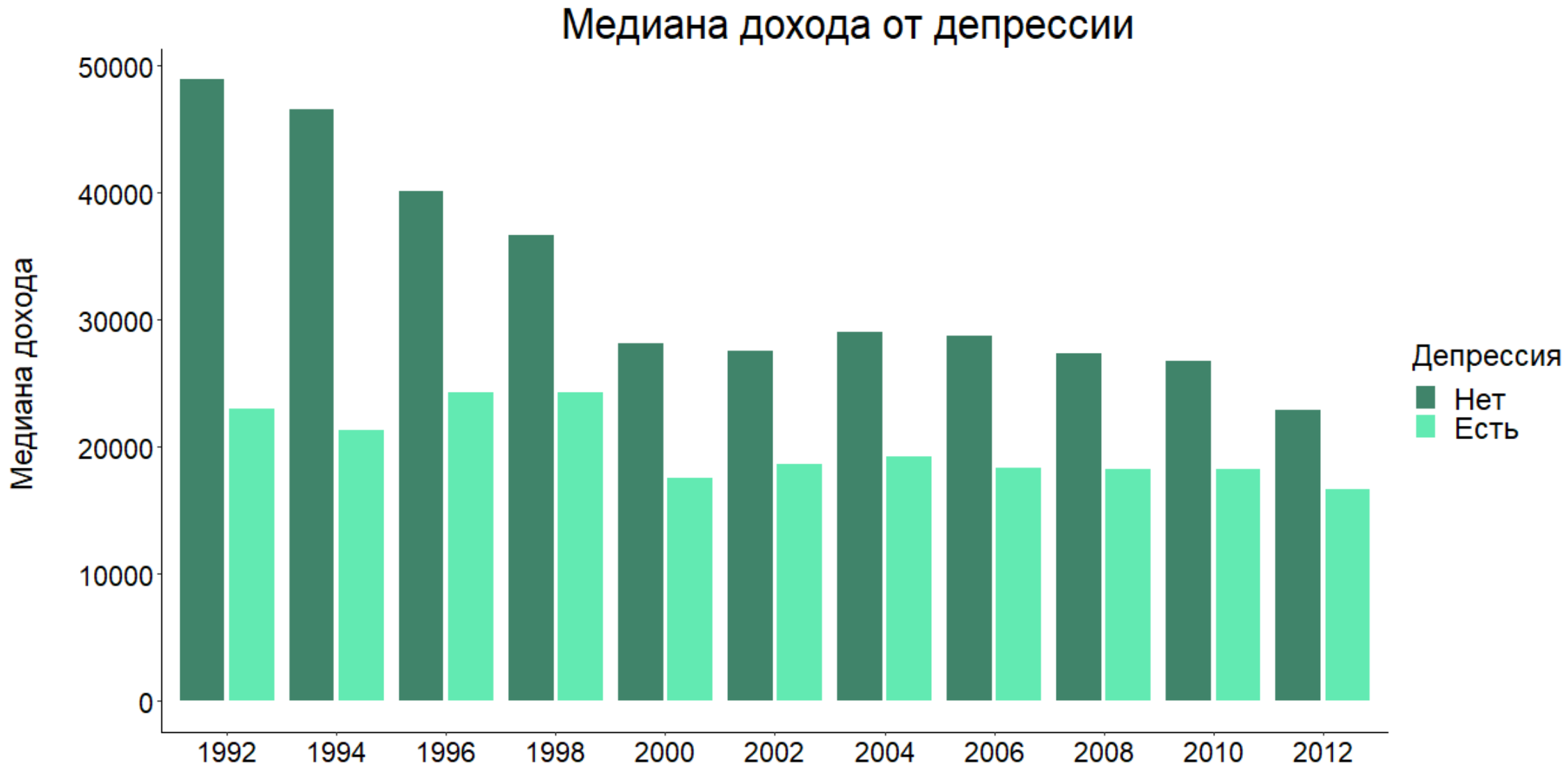
Выборка

- Объём: 170 тыс. интервью
- Респонденты от 50 до 90 лет
- Респонденты разделены на когорты по поколениям

Генеральная совокупность

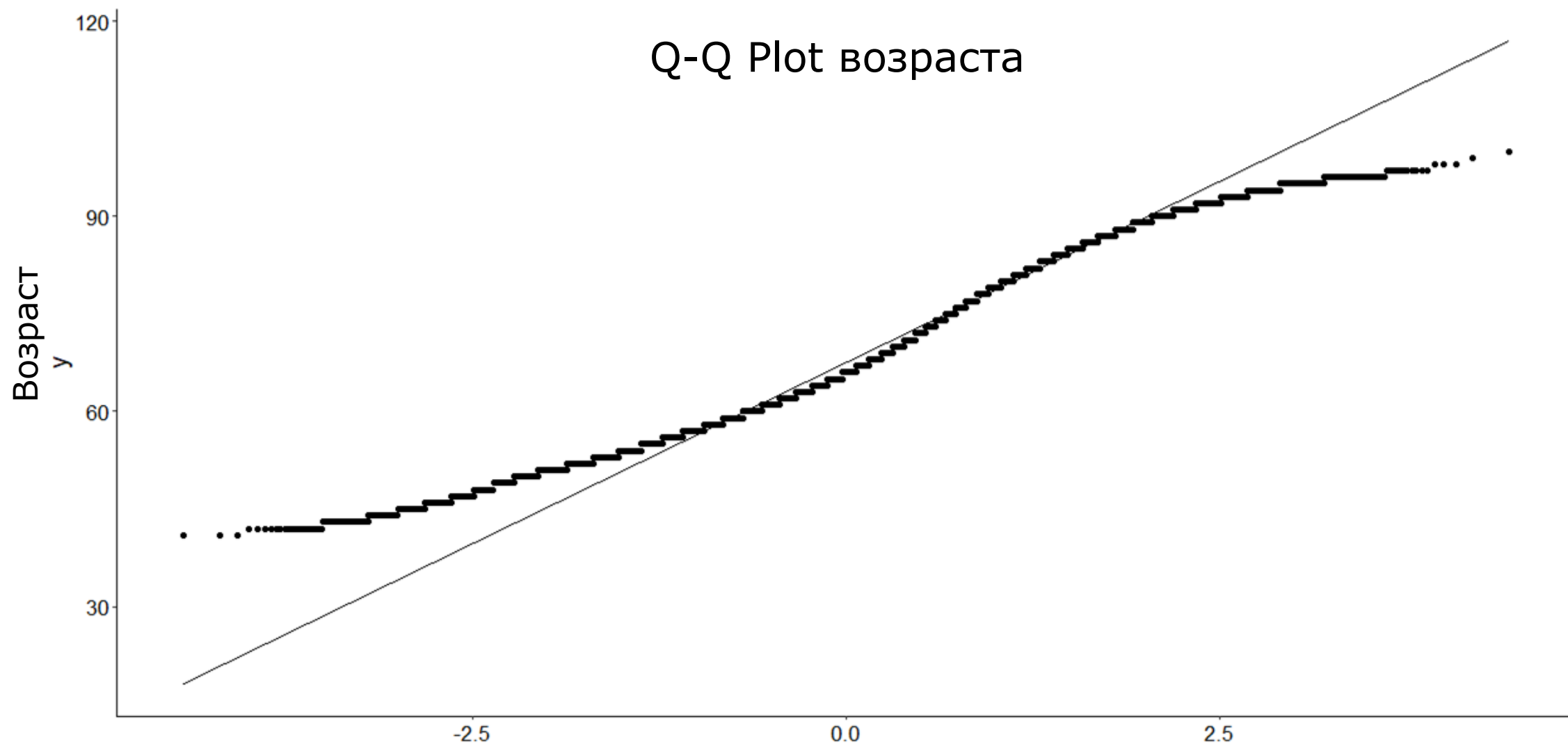
- Жители различных американских штатов
- Преклонного возраста
- Различных поколений
- В период с 1991 по 2012 год

У людей без депрессии доход выше

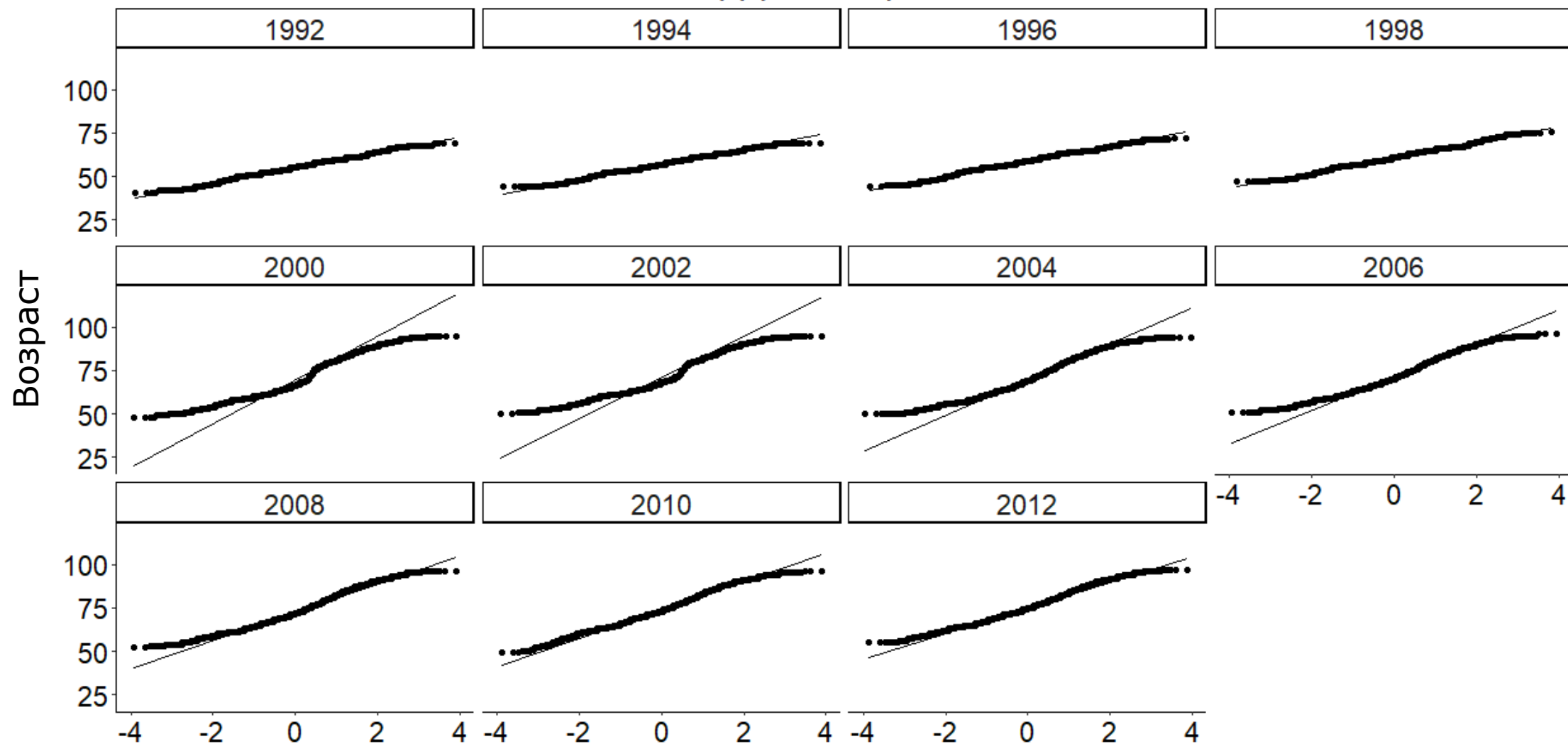


Q-Q Plot возраста

Q-Q Plot возраста



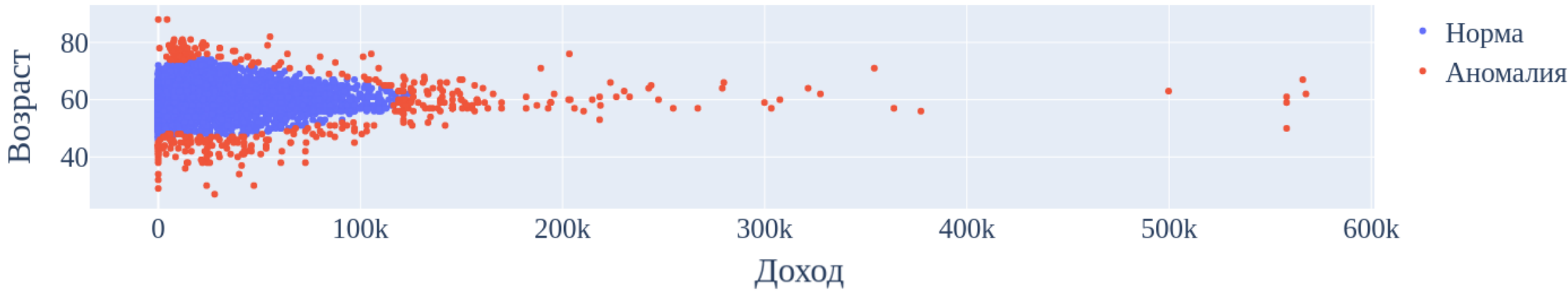
Q-Q Plot возраста по годам



Была использована модель Isolation forest

Isolation forest

График модели за 2003-2004 год



Результаты

Год	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012
Всего данных	10397	8888	8551	8256	11981	10923	13642	12673	11724	10477	9721
Аномалии	393	393	373	373	469	442	562	534	507	465	436

Пропуски значений

В период 1991-1992 более
половины данных – пропуски по
зависимой переменной

До 2000 пропусков по зависимой
переменной за каждую прочую волну

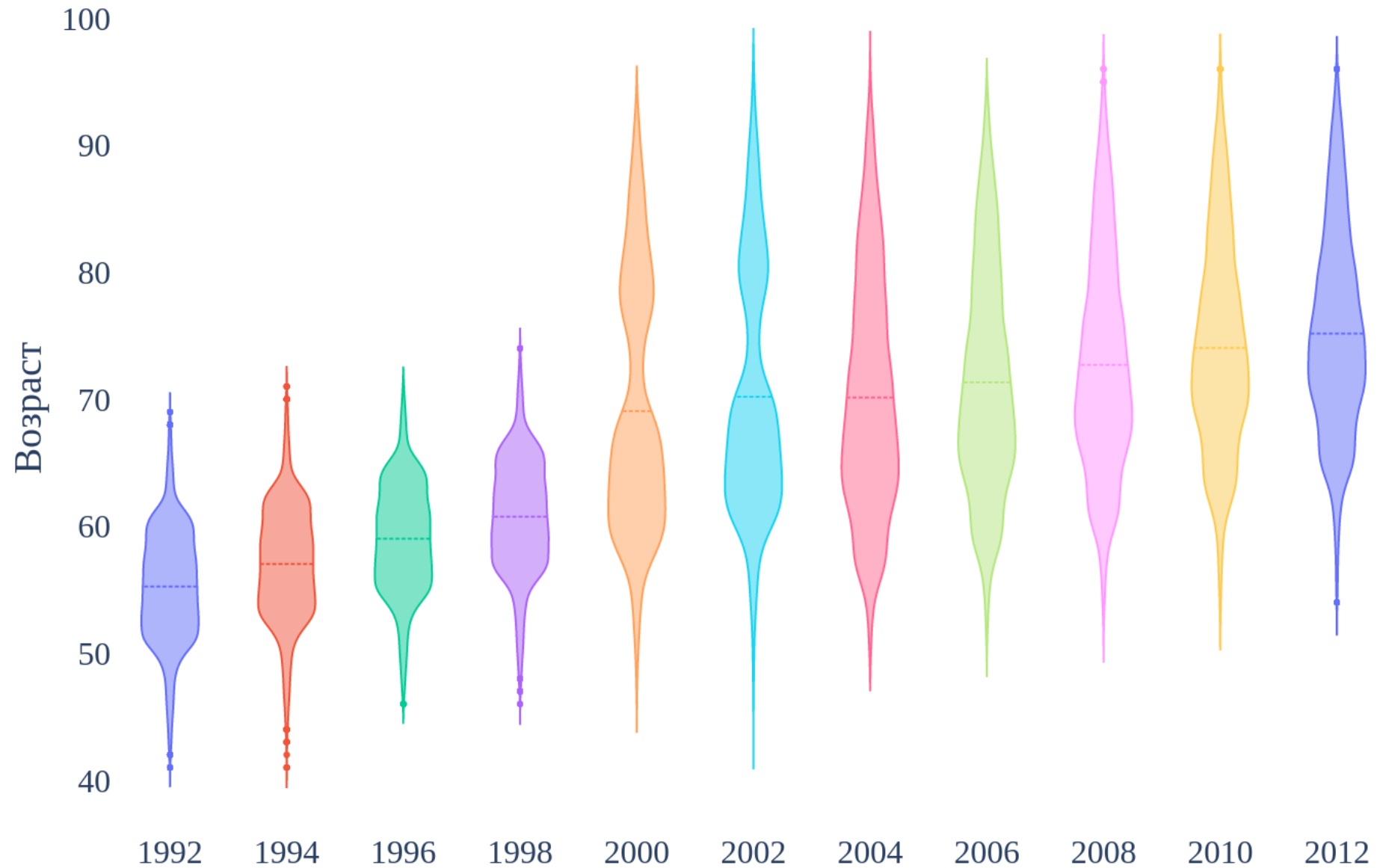
87 пропусков в семейном
положении

26 пропусков в данных о
количестве лет обучения

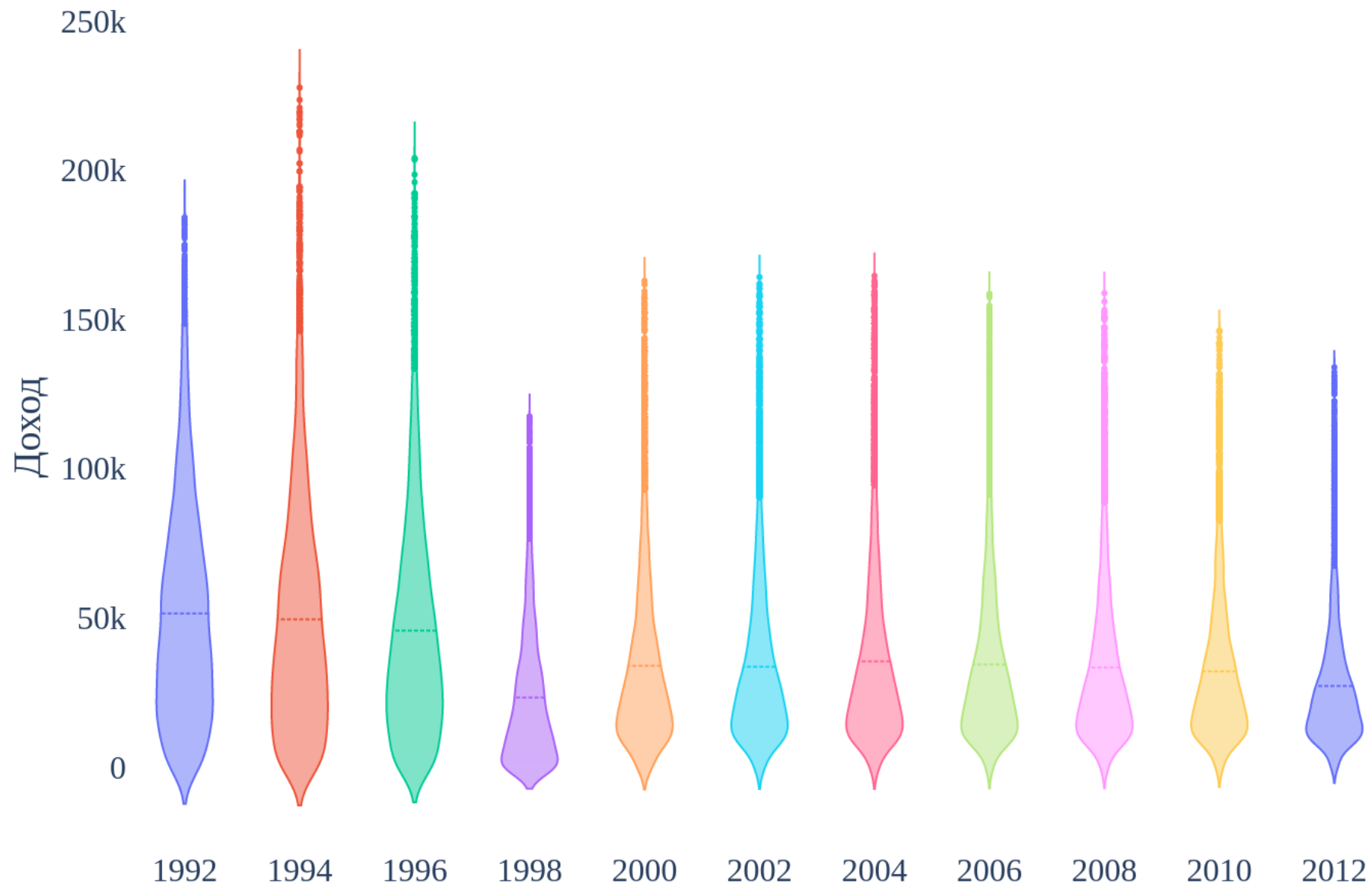


Удаление
записей

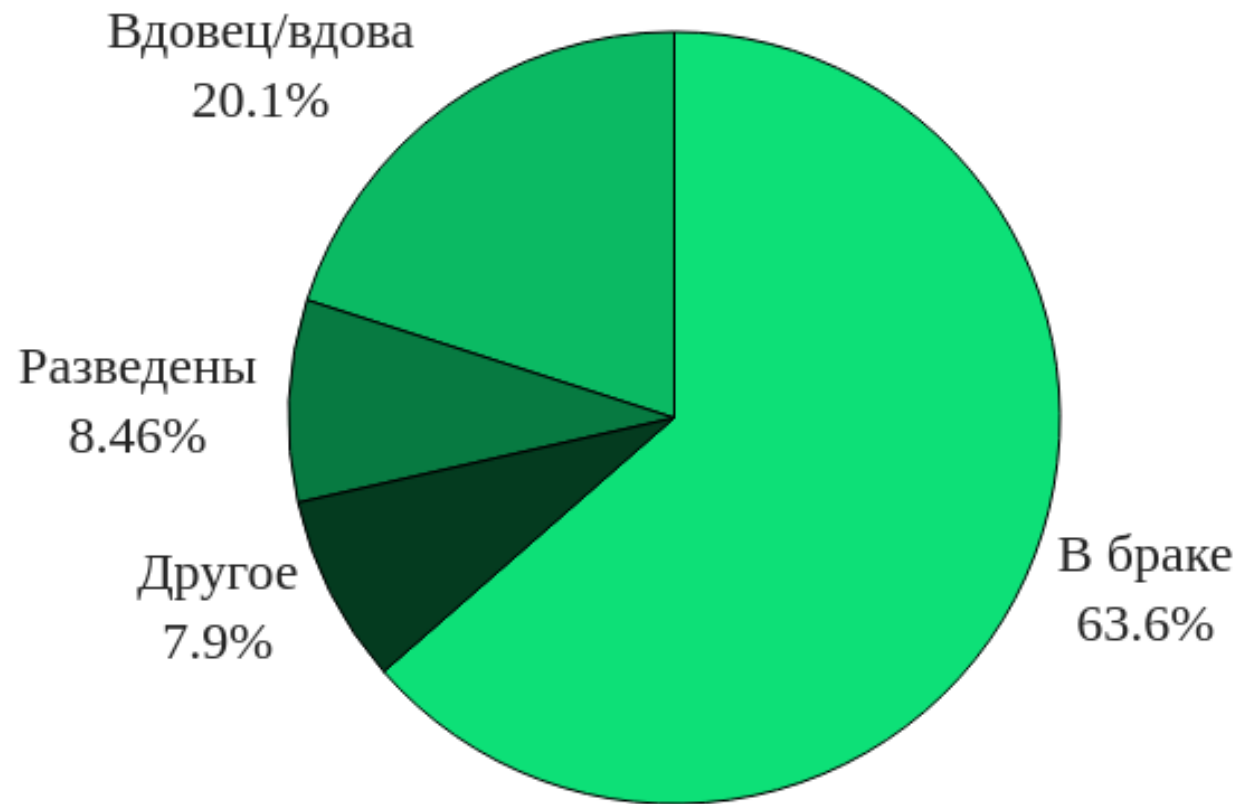
Распределение возраста респондентов по годам опроса



Распределение дохода респондентов по годам опроса



Распределение значений у переменной "семейное положение"



Дизайн исследования

Разбиение на года

Исключение от автокорреляции по времени
Исключение автокорреляции по человеку

Логистическая регрессия
с добавлением факторов

Уменьшение смещения благодаря
включению значимых факторов

Дизайн исследования



Делаем данные более полезными

$$depres = (CES_D_8 \geq 3)^{[3]}$$

family_issues ~ если семейное положение среди:

- ✓ Вдова/вдовец
- ✓ Разведены
- и др.

$$retired = (Статус\ на\ рынке\ труда == на\ пенсии)$$

$$\log_income = \log(hhEarn + hhPension + hhPayments + 1)$$

Дополнительные
данные

Количество
лет обучения

Опрос по
CES-D 8

На зависимую переменную влияет ряд прочих факторов

Характеристика	Корреляция со стрессом	Корреляция с доходом
Проблемы в семье	Источник стресса	Могут сказаться на эффективности
Длительность обучения	Бóльшая толерантность к стрессу	Более высокая квалификация
Возраст	Осознание старения	Больше опыта
Пол	Различная физиология и социальное положение	“Стеклянный потолок”

На зависимую переменную влияет ряд прочих факторов

Характеристика	Корреляция со стрессом	Корреляция с доходом
На пенсии (не работает)	Разрушение карьерного целеполагания	Не получает зарплату
Инвалид не на пенсии	Социальная изолированность	Получает пособие
Безработный (не инвалид, не на пенсии)	Возможно падение самооценки, нарушение социализации	Не получает зарплату

Уравнение модели

$$\begin{aligned} \textit{depres} \sim & \beta_0 + \beta_1 * \textit{log_income} + \beta_2 * \textit{log_income}^2 + \\ & + \beta_3 * \textit{family_issues} + \beta_4 * \textit{gender} + \beta_5 * \textit{age} + \\ & + \beta_6 * \textit{retired} + \beta_7 * \textit{unemployed} + \\ & + \beta_8 * \textit{disabled} + \beta_9 * \textit{study_duration} + \varepsilon_i \end{aligned}$$

Рассматриваемый коэффициент

Год	Log_income	Log_income ²	Pseudo-R ²	AUC ROC
1991-1992	0.343 ***	-0.034 ***	0.121	0.739
1993-1994	0.345 ***	-0.035 ***	0.126	0.737
1995-1996	0.148 ***	-0.015 ***	0.104	0.718
1997-1998	0.180 ***	-0.016 ***	0.080	0.693
1999-2000	0.227 ***	-0.023 ***	0.077	0.690
2001-2002	0.244 ***	-0.025 ***	0.070	0.686
2003-2004	0.086 *	-0.011 ***	0.077	0.690
2005-2006	0.173 ***	-0.019 ***	0.087	0.704
2007-2008	0.236 ***	-0.023 ***	0.067	0.680
2009-2010	0.155 ***	-0.017 ***	0.067	0.682
2009-2012	0.091	-0.013 ***	0.062	0.672

Был проведён VIF тест

Log income² хорошо предсказывается другими предикторами

Возникает сильное смещение
оценки коэффициента

Исключение этой переменной не
сильно скажется на результате

Год	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012
VIF	16.94	18.06	19.31	17.06	16.33	14.77	13.91	13.33	12.79	12.57	12.19

Уравнение модели

$$\begin{aligned} \text{depres} \sim & \beta_0 + \beta_1 * \text{log_income} + \beta_2 * \text{log_income}^2 + \\ & + \beta_3 * \text{family_issues} + \beta_4 * \text{gender} + \beta_5 * \text{age} + \\ & + \beta_6 * \text{retired} + \beta_7 * \text{unemployed} + \\ & + \beta_8 * \text{disabled} + \beta_9 * \text{study_duration} + \varepsilon_i \end{aligned}$$

Коэффициенты преобразованной модели

Год	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012
<i>Log_income</i>	-0.04	-0.04	-0.02	-0.01	-0.03	-0.04	-0.04	-0.05	-0.04	-0.05	-0.06
Значимость	***	***			***	***	***	***	**	***	***

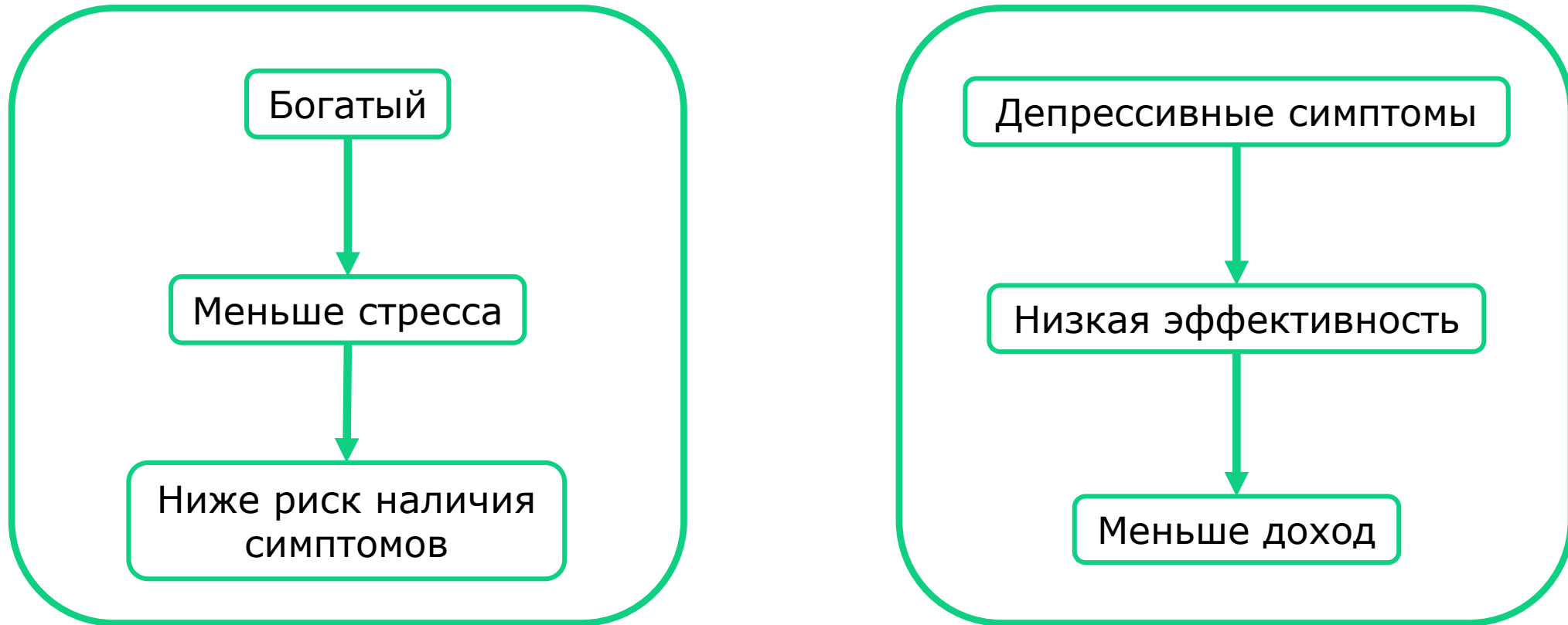
Предполагается смещение оценок коэффициентов переменной интереса из-за эндогенности введенной двусторонней причинно-следственной взаимосвязью^[4-5]

[4] Ольга Анатольевна Демидова & Дмитрий Игоревич Малахов (2017). "Эконометрика"

[5] Филипп Картаев (2022). Введение в эконометрику.

Мотивация использовать инструментальные переменные

Двусторонняя причинно-следственная взаимосвязь



Инструментальные переменные

**Вспомогательная регрессия (IV)
на основе уравнения Минсера^[7]**

$$\widehat{income} = \pi_0 + \pi_1 * study_duration + \\ + \pi_2 * age + \pi_3 * age^2 + \pi_4 * retired + \\ + \pi_5 * disabled + \pi_6 * unemployed + \varepsilon_i$$

[7] Jacob Mincer (1958). "Investment in human capital and personal income distribution". In: *Journal of political economy* 66.4, pp. 281–302.

Формула Logit (IV)

$$\begin{aligned} \textit{depres} \sim & \beta_0 + \beta_1 * \log(\widehat{\textit{income}}) + \\ & + \beta_2 * \textit{family_issues} + \beta_3 * \textit{gender} + \varepsilon_i \end{aligned}$$

Логит (IV)

Год	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012
<i>Log_income</i>	-0.85	-1.07	-1.09	-0.87	-0.5	-0.45	-0.73	-0.66	-0.61	-0.64	-1.01
Значимость	***	***	***	***	***	***	***	***	***	***	***

Выводы

Нет оснований, чтобы отклонить нулевую гипотезу о том, что логарифм дохода значимо влияет на вероятность наличия симптомов депрессии

Нет оснований, чтобы отклонить гипотезу об обратной взаимосвязи между логарифмом дохода и вероятностью наличия депрессивных симптомов

Влияние дохода на наличие депрессивных симптомов не изменилось за годы опроса

$$\text{depres} \sim \text{year}_i * (\beta_0 + \beta_1 * \text{family_issues} + \beta_2 * \text{gender}) + \\ + \text{year}_i * \beta_3 * \log(\widehat{\text{income}}) + \varepsilon_i$$

?
=

$$\text{depres} \sim \text{year}_i * (\beta_0 + \beta_1 * \text{family_issues} + \beta_2 * \text{gender}) + \\ + \text{year}_i * \beta_3 * \log(\widehat{\text{income}}) + \varepsilon_i$$

Результат сравнение моделей

$$\chi^2 : 167.3 ***$$

Достаточно оснований отклонить нулевую гипотезу о том, что взаимодействие года и дохода домохозяйства незначимо влияет на вероятность наличия депрессивных симптомов

Достаточно оснований полагать, что влияние дохода на наличие депрессивных симптомов значимо отличается в различные годы опроса

Достаточно оснований отклонить вторую гипотезу

Результат сравнение моделей

Перспективы:

- Уменьшить смещение, добавив важные переменные (например потребление алкоголя и занятия спортом)
- Распространение генеральной совокупности на большее число периодов / территорий / поколений
- Проведение когортного анализа по поколениям

Ограничения:

- Смещение оценок из-за невключения ненаблюдаемых переменных, формата сбора данных, самоотбора
- Замена латентной переменной (наличие депрессии) измеряемой
- Наличие внешних фиксированных факторов
- Автокорреляция по времени и по респонденту
- Чёткие ограничения генеральной совокупности по территории, времени и поколениям

Результат сравнение моделей

Практическая польза:

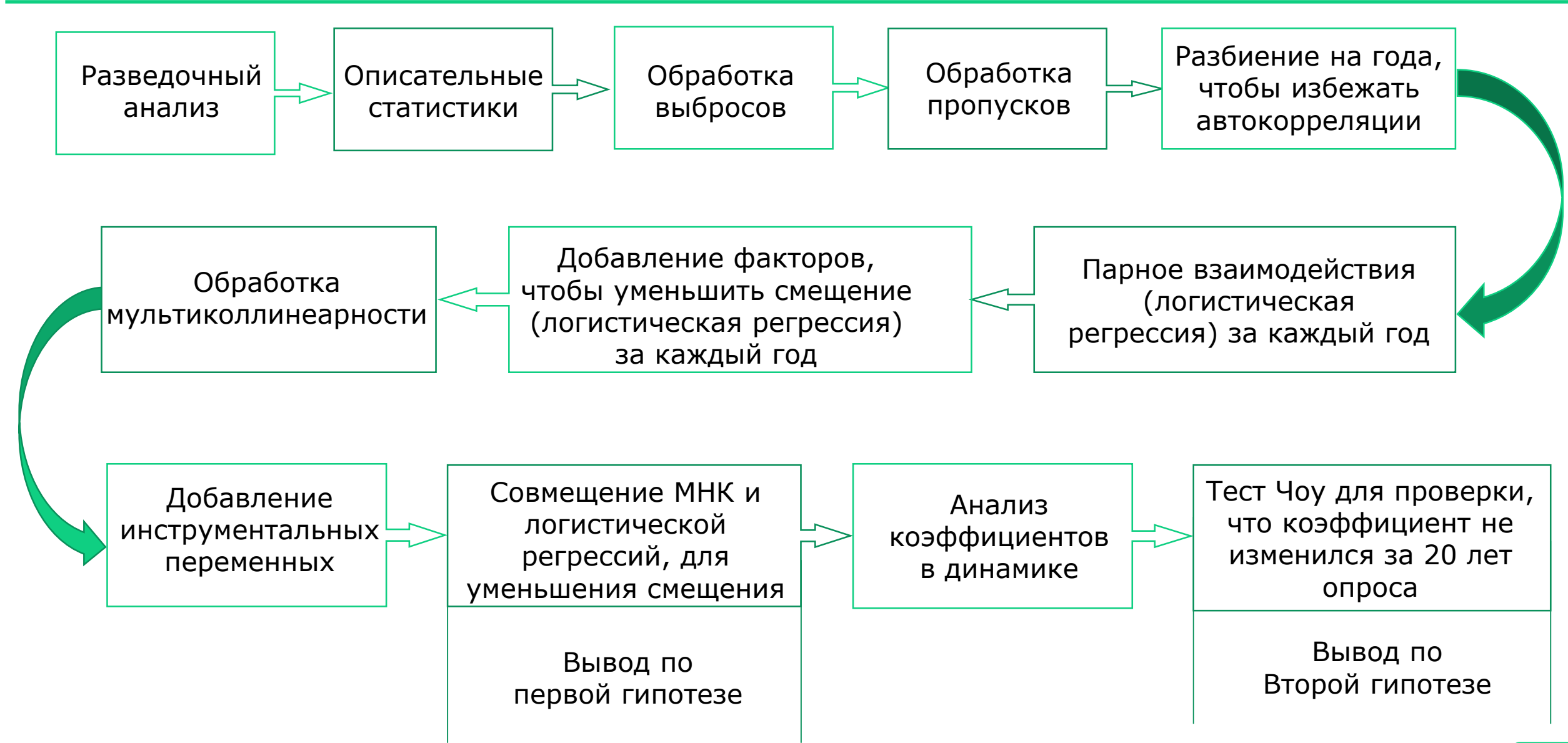
- адаптация социально-экономических программ предотвращения ментальных расстройств и поддержки больных ими
- Адаптация зарплатной политики для должностей требующих высокой квалификации и опыта

policy implication

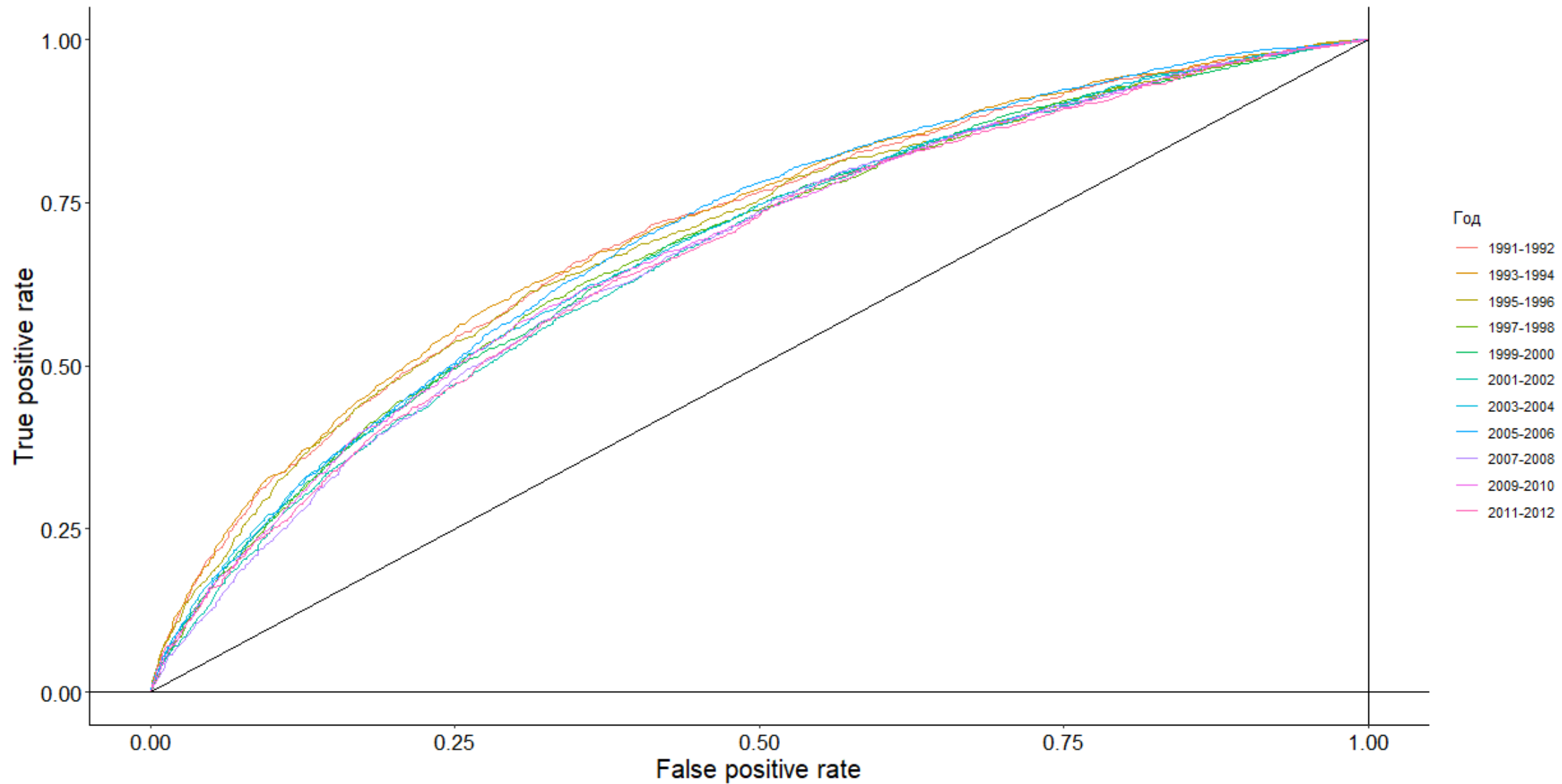
Повышение заработной платы работника до оптимума позволяет максимизировать прибыль, снизив риск потери эффективности

ПРИЛОЖЕНИЕ

Аккуратно обходим ограничения датасета



ROC кривые по годам



**Корреляция предсказанных и
изначальных значений (IV)**

Год	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012
Корреляция	50.8%	48.4%	46.4%	46.2%	46.3%	44.6%	46.1%	43%	40.5%	37%	36.9%