

# О доходе и рождаемости

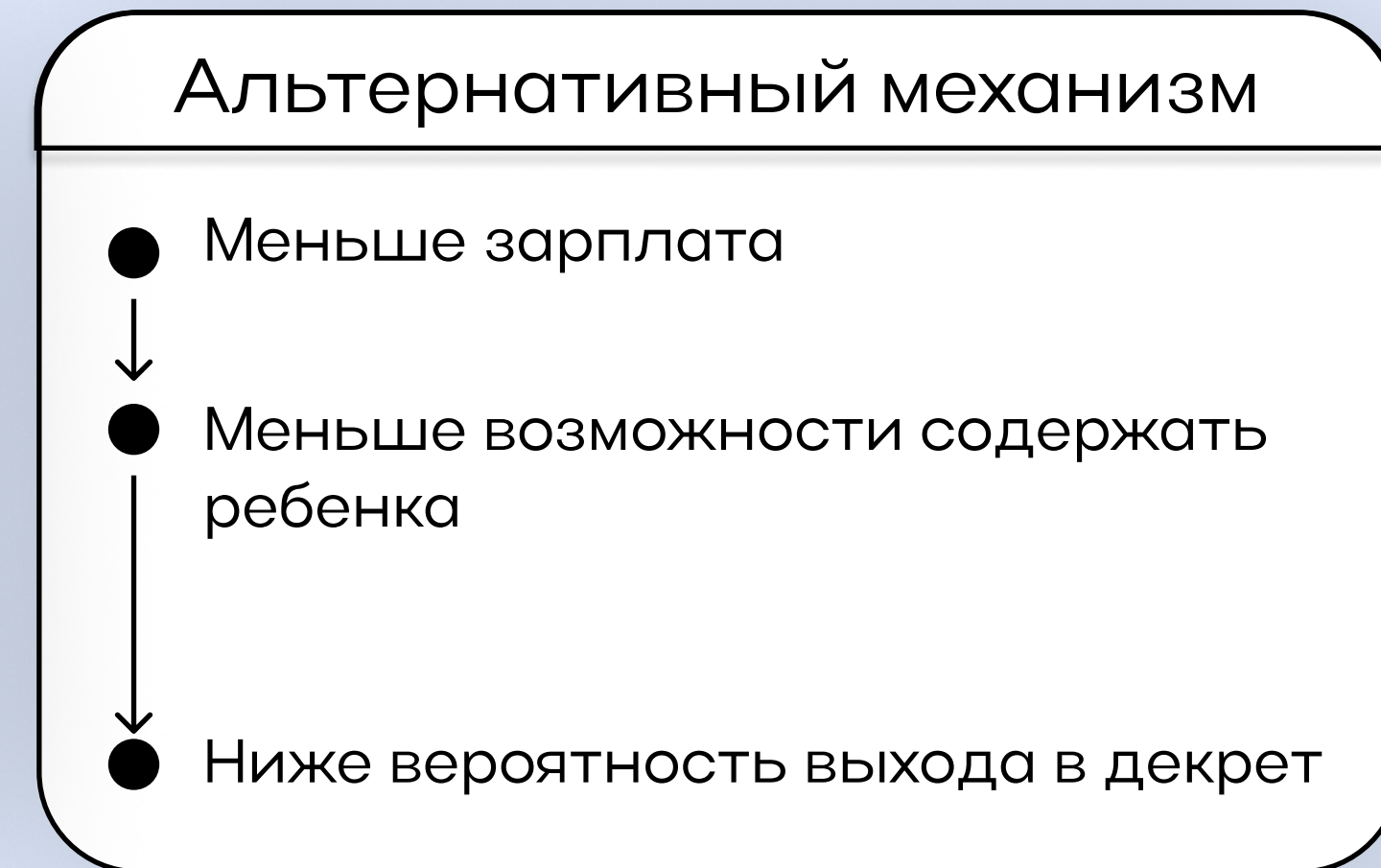
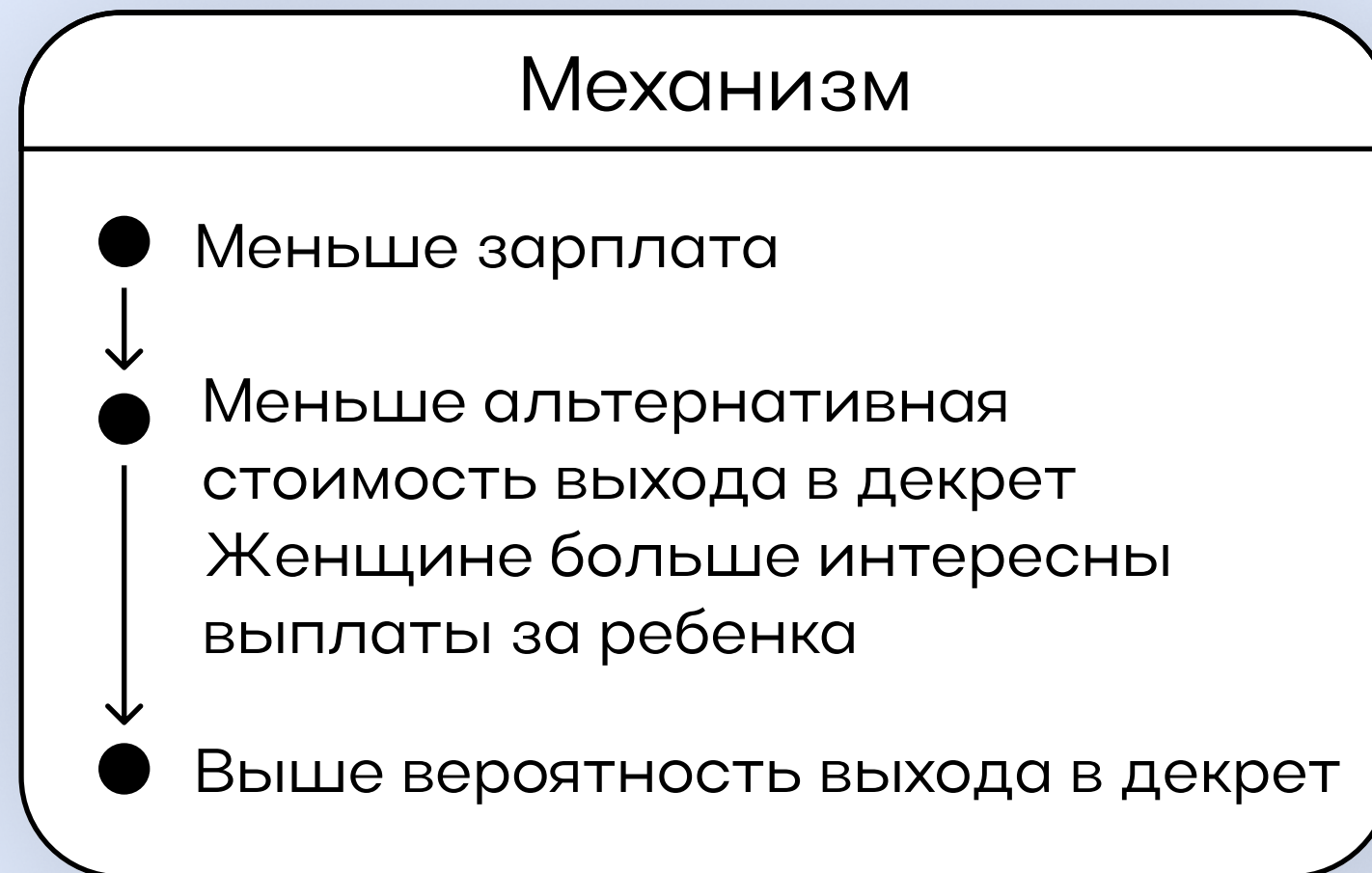
Исследование по потребительским финансам Банка России

Бакалова Екатерина  
Карташов Игорь  
Кириченко Дмитрий  
Иванов Герман

# Исследовательский вопрос

Как доход женщины влияет на решение родить ребенка?

## Механизм



## Гипотеза

Женщина с меньшим доходом вероятнее решит родить ребенка



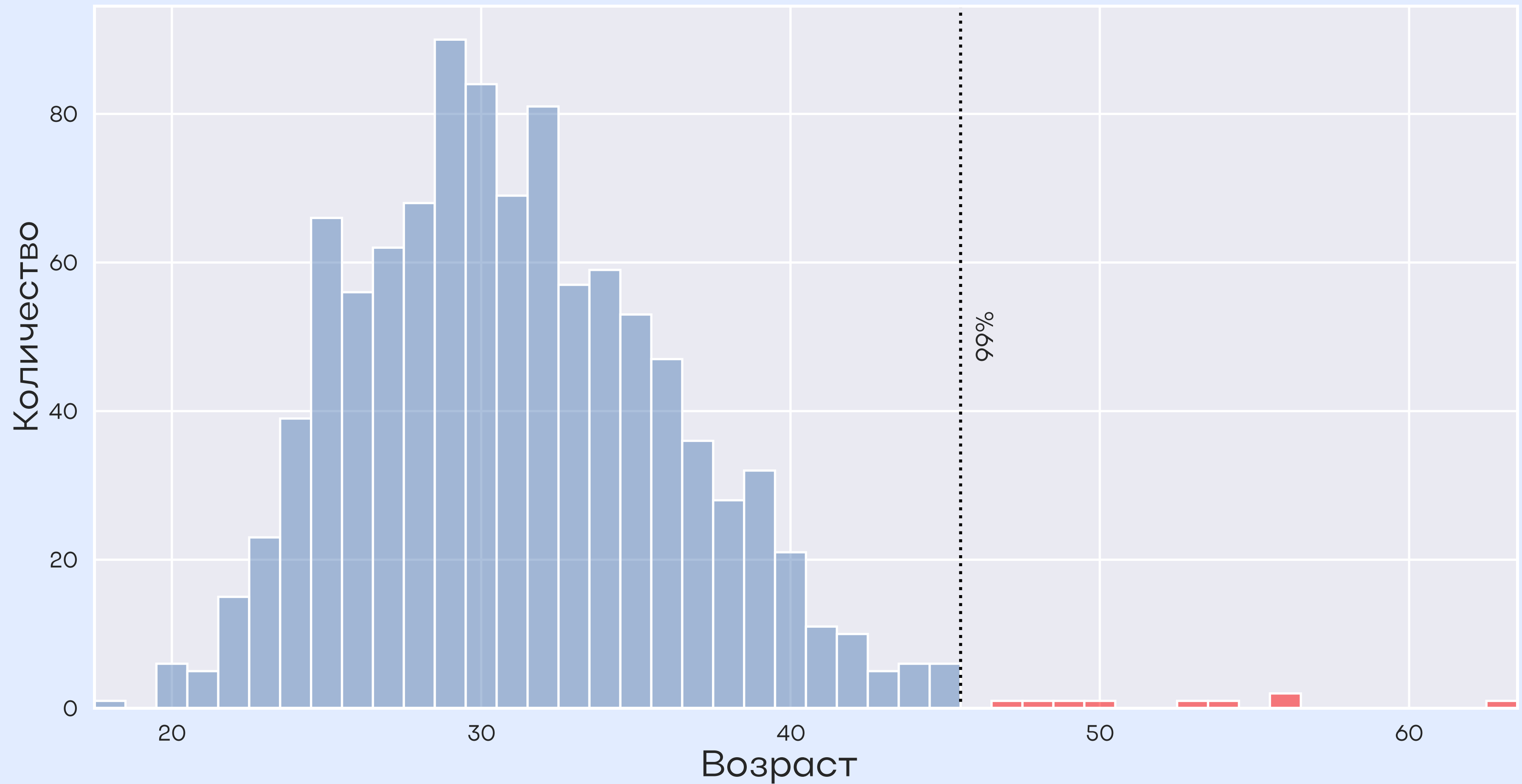
## Структура данных

- Опрос членов домохозяйств по отдельности и домохозяйств в целом (разные опросники)
- Данные за 2013, 2015, 2018, 2020, 2022 года.  
Сплит-панель.
- Вопросы о финансовом и материальном положении, мнении о экономической ситуации
- Интервьюирование с последующей выборочной проверкой

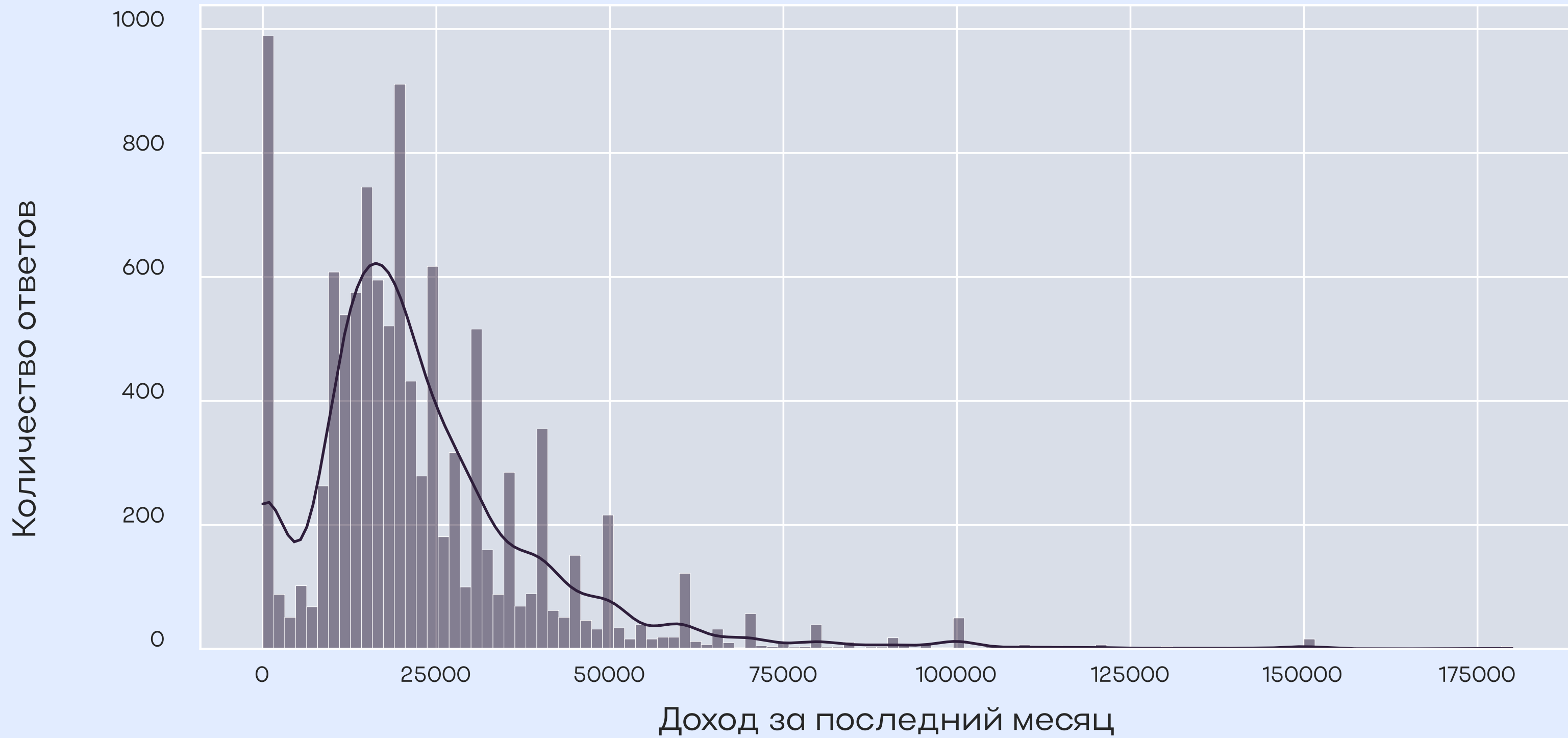
## Формирование выборки в опросе

- Формирование выборки по адресам жилищ
- Выборка репрезентативная для населения России
- Но нельзя сравнивать регионы и не представлен верхний квантиль по доходу

Распределение возрастов среди женщин в декрете

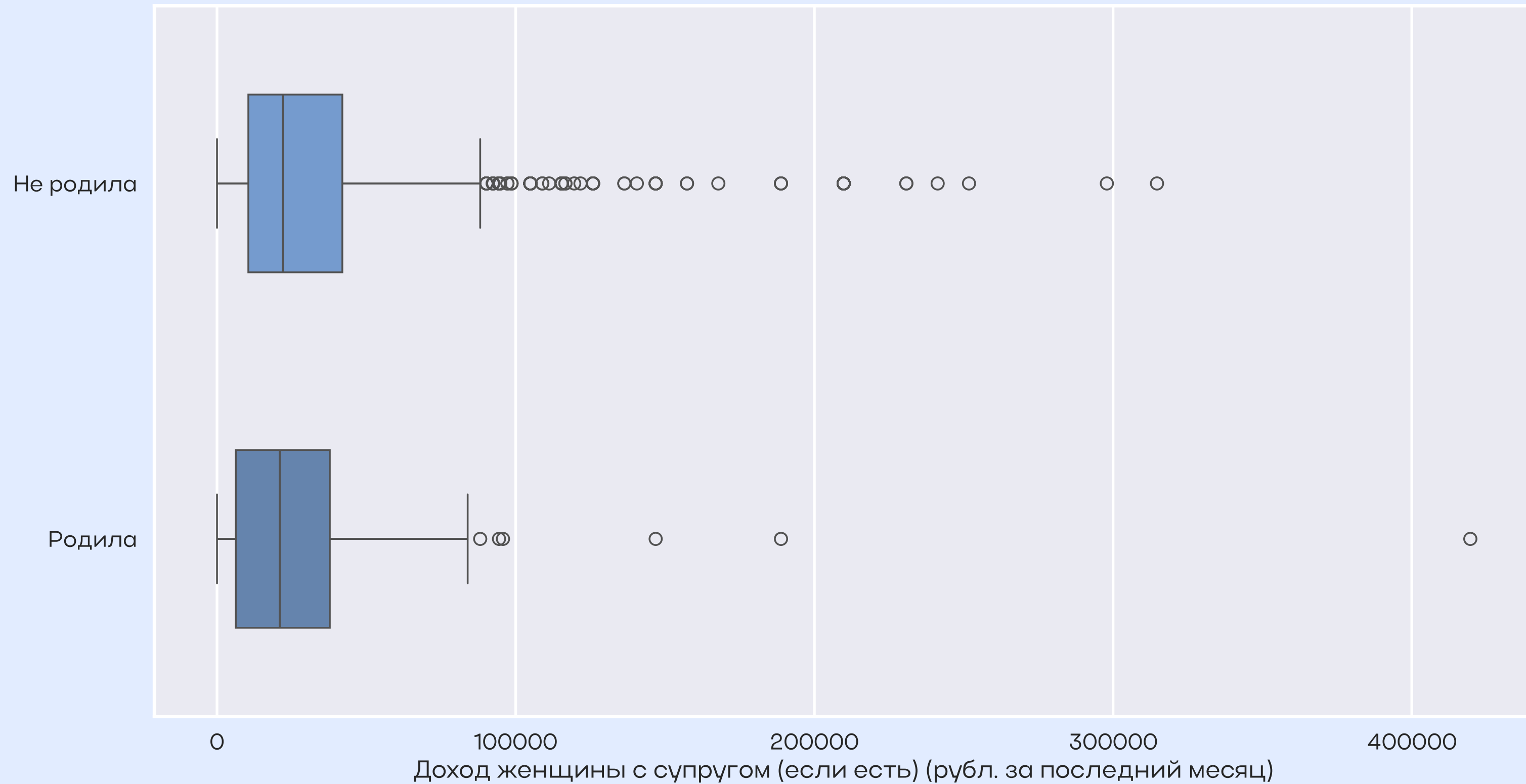


### Распределение доходов респондентов в 2020 г. после удаления выбросов



## Зависимость зарплаты и родов

Распределение общих доходов женщин в 2020 г.





Создание новой переменной - **реальный доход**

Дает возможность сравнивать разные года

Посчитанно с помощью значения ИПЦ с Росстата

Распределение доходов респондентов разных волн



# Логистическая регрессия

Сравниваем года попарно

В этих годах мы  
смотрим  
переменные  
предикторы

**2013** —————• **2015**  
**2015** —————• **2018**  
**2018** —————• **2020**  
**2020** —————• **2022**

В этих годах  
зависимая  
переменная:  
выйдет женщина  
в декрет или нет

# Логистическая регрессия

Сравниваем года попарно

Переменные-предикторы:

- возраст
- качество отношений с мужем
- уровень образования
- социальные льготы от работодателя
- уверенность в том, что женщина сможет сменить работу
- материальное положение
- мнение о подходящем моменте сбережений
- **общий доход (логарифмированный и индексированный)**

**2013** ————— • **2015**  
**2015** ————— • **2018**  
**2018** ————— • **2020**  
**2020** ————— • **2022**

В этих годах  
зависимая  
переменная:  
выйдет в декрет  
или нет

Замолняем модами значения, где нет ответа

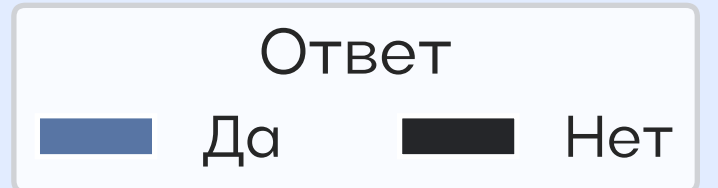




# Логистическая регрессия

Распределение дополнительных предикторов

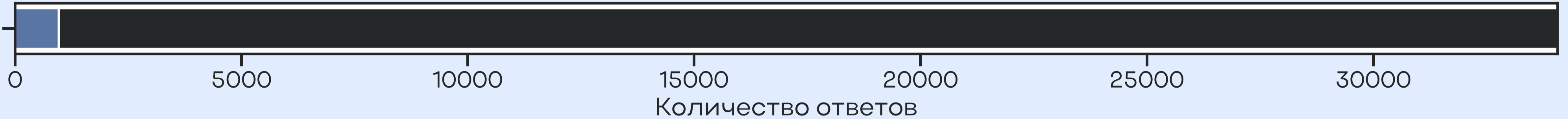
Социальные льготы, предоставляемые работодателем



Оплата отпуска по беременности, родам, по уходу за ребенком



Бесплатное содержание детей в дошкольных учреждениях или его частичная оплата

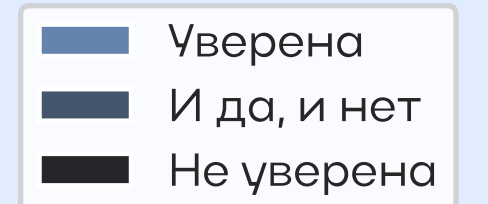
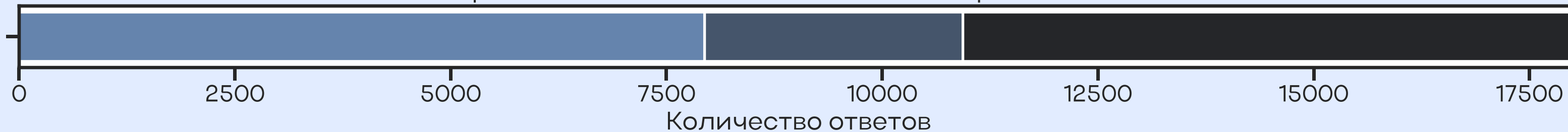


# Логистическая регрессия

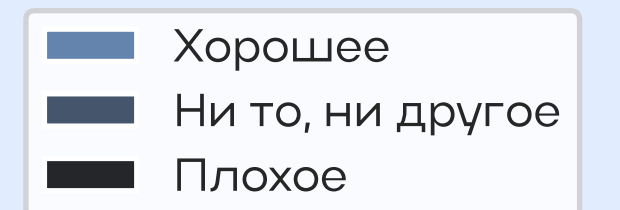
## Распределение дополнительных предикторов

Воспринимаемый уровень экономической стабильности женщин

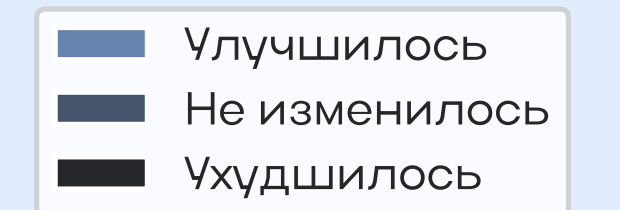
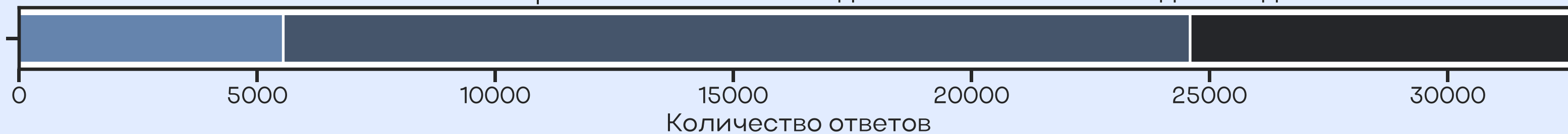
Уверенность в возможности найти новое рабочее место



Мнение относительно безопасности сбережений



Изменение материального положения домохозяйства за последний год



# Логистическая регрессия

## Проверка предпосылок модели

Матрица корреляций между факторами

income\_woman\_and\_husband

0.20	0.02	-0.00	0.02	-0.02	-0.11	-0.53	0.05	0.06	0.15	0.01	0.27	0.13
age	household_fin_sit_better	household_fin_sit_same	time_for_save_good	time_for_save_neither	paid_matrnity_leave	free_daycare	re_satisfaction	middle_school	college	university	job_demand_conf	job_demand_unsure

- Предпосылка регрессии — отсутствие мультиколлинеарности
- Строгой мультиколлинеарности нет
- Частичная коллинеарность есть, но:
  - не нарушает предпосылки
  - не влияет на сходимость
  - не смещает оценки коэффициентов

[1]

# Логистическая регрессия

$H_0$

Доход не значимо влияет на вероятность решения родить ребенка

$H_1$

Доход значимо влияет на вероятность решения родить ребенка



# Логистическая регрессия

$$\begin{aligned} child \sim & \beta_1 + \beta_2 \cdot year\_2015 + \beta_3 \cdot year\_2018 + \beta_4 \cdot year\_2020 + \beta_5 \cdot age + \beta_6 \cdot \\ & household\_financial\_situation\_better + \beta_7 \cdot household\_financial\_situation\_same + \\ & \beta_8 \cdot time\_for\_savings\_good + \beta_9 \cdot time\_for\_savings\_neither + \\ & \beta_{10} \cdot \ln(income\_woman\_and\_husband) + \beta_{11} \cdot paid\_maternity\_leave + \\ & \beta_{12} \cdot free\_healthcare + \beta_{13} \cdot free\_daycare + \beta_{14} \cdot relationship\_satisfaction + \\ & \beta_{15} \cdot middle\_school + \beta_{16} \cdot college + \beta_{17} \cdot university + \beta_{18} \cdot job\_demand\_confident + \\ & \beta_{19} \cdot job\_demand\_unsure \end{aligned}$$

# Логистическая регрессия

По результатам логистической регрессии переменная года оказалась незначима, а следовательно мы можем сравнивать все волны опроса в одной модели, и не учитывать переменную года

Переменная	P-value
year_2015	0.469
year_2018	0.058
year_2020	0.398

# Логистическая регрессия

$$\begin{aligned} child \sim & \beta_1 + \beta_2 \cdot \cancel{year\_2015} + \beta_3 \cdot \cancel{year\_2018} + \beta_4 \cdot \cancel{year\_2020} + \beta_5 \cdot age + \beta_6 \cdot \\ & household\_financial\_situation\_better + \beta_7 \cdot household\_financial\_situation\_same + \\ & \beta_8 \cdot time\_for\_savings\_good + \beta_9 \cdot time\_for\_savings\_neither + \\ & \beta_{10} \cdot \ln(income\_woman\_and\_husband) + \beta_{11} \cdot paid\_maternity\_leave + \\ & \beta_{12} \cdot free\_healthcare + \beta_{13} \cdot free\_daycare + \beta_{14} \cdot relationship\_satisfaction + \\ & \beta_{15} \cdot middle\_school + \beta_{16} \cdot college + \beta_{17} \cdot university + \beta_{18} \cdot job\_demand\_confident + \\ & \beta_{19} \cdot job\_demand\_unsure \end{aligned}$$

# Интерпретация результата

Переменная	Coef	P-value
income_woman_and_husband	-0.1165	0.000



## Интерпритация результата

Зарплата	Логарифм зарпоаты	Коэффициент	Вероятность
22026	10	$10 * -0.1165 = -1,165$	x%
36315	10,5	$10.5 * -0.1165 = -1,2233$	x - 1.4%
59874	11	$11 * -0.1165 = -1.2815$	x - 2.8%

**\*при прочих равных**

$$\hat{P} = \frac{1}{1 + e^{-(0.8399 - 0.1165 \cdot \text{income\_woman\_and\_husband})}}$$

## Проверка модели на подвыборках

Год	Coef	P-value	
2015	-0.1267	0.000	
2018	-0.1399	0.000	
2020	-0.1917	0.000	

# Ограничения → Перспективы

Из-за срезов годов  
невозможно отследить все  
факты выхода в декрет/  
замуж

Расширить опросник,  
включить дату начала  
интересующих нас  
переменных

Не представлен верхний  
квантиль населения по  
доходу, а для них могут  
работать другие  
закономерности

Изменить дизайн  
исследования и  
включить эту страту



## Практическая польза

Для государства:

Правильно проводить свою демографическую политику

Для фирм:

Грамотно выстраивать зарплатную политику для своих сотрудниц

## Policy Implication

- Повышение процента выплат по уходу за ребенком
- ↓
- Меньшая альтернативная стоимость рождения ребенка
- ↓
- Больше количество женщин, выходящих в декрет

- Повышение зарплаты
- ↓
- Для них альтернативная стоимость рождения ребенка повышается
- ↓
- Сохранение в штате большего количества квалифицированных сотрудниц



# Правило ~~пяти~~ четырех сигм



**Герман**

Предварительный анализ,  
визуализация



**Екатерина**

Тимлид,  
мат модель



**Игорь**

Реализация мат. модели



**Дмитрий**

Ресерч, программист





[Ссылка на Google Диск](#)



[Ссылка на Google Диск](#)





[Ссылка на Google Диск](#)



[Ссылка на Google Диск](#)