



Федеральное государственное автономное образовательное учреждение высшего
образования

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА
ЭКОНОМИКИ»

Факультет экономических наук, образовательная программа «Экономика»

ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ЗАРАБОТНОЙ ПЛАТЫ ОТ РАЗЛИЧНЫХ
ФАКТОРОВ НА РОССИЙСКОМ РЫНКЕ ТРУДА

Выполнили студентки группы БЭК217:

Кудабаева Гульшат Вильдановна

Нежурина Екатерина Владиславовна

Уначева Амина Аслангериевна

Преподаватель:

Демьяненко Артем Владимирович

Москва 2024

Оглавление

Введение.....	3
Данные и переменные.....	4
Итоговый набор переменных:.....	5
Проблема мультиколлинеарности.....	6
Проблема эндогенности.....	6
Предварительный анализ данных.....	7
Модель.....	11
Интерпретация результатов.....	14
Список литературы.....	16
Приложение.....	17

Введение

В настоящее время динамика заработной платы играет ключевую роль в формировании социально-экономических отношений и обеспечении жизненного уровня населения. Однако, разнообразие факторов, влияющих на этот показатель, создает сложную и многогранную картину, требующую внимательного анализа и исследования. В контексте российского рынка труда этот вопрос становится особенно актуальным, учитывая динамику экономических процессов, политические и социокультурные изменения.

В научных исследованиях значительная дифференциация в заработной плате между возрастными группами объясняется как обесцениванием человеческого капитала работника в конце его трудовой карьеры, так и влиянием возрастной дискриминации [1]. Определение причин дифференциации в уровне заработной платы представляет непростую задачу. В нашем исследовании мы проверим, как человеческий капитал влияет на заработную плату, проверив следующую гипотезу:

Гипотеза 1. Заработная плата индивидов положительно зависит от качества человеческого капитала (образования и здоровья).

В настоящее время вопросы семейной политики и социальной защиты являются одними из ключевых аспектов обсуждений в контексте экономического развития и социального благосостояния. Одним из интересующих аспектов в этой области является влияние наличия детей на заработную плату работников женского пола. В современных исследованиях уже доказано наличие отрицательной корреляции между уровнем заработной платы женщины и наличием у нее детей до 18 лет [2][3]. В связи с этим, мы хотим проверить, подтвердится ли следующая гипотеза для российского рынка труда.

Гипотеза 2. Наличие детей до 18 лет отрицательно влияет на заработную плату женщин, то есть, чем больше детей, тем меньше уровень заработной платы.

Британские ученые провели исследование и выяснили, что высокие мужчины имеют более высокий заработок [4]. Нам стало интересно :) , окажется ли рост респондентов значимым фактором, определяющим заработную плату. Следовательно, чтобы проверить выполняется ли данная закономерность на российском рынке труда, мы выдвинули последнюю гипотезу.

Гипотеза 3. Рост мужчины положительно влияет на размер заработной платы. Предполагаемая зависимость может быть связана с преимуществами, которыми обладают высокие люди: возможно уверенность в себе, наличие лидерских качеств или физическая выносливость.

Данные и переменные

В качестве основного источника данных для исследования факторов, влияющих на формирование заработной платы на российском рынке труда, мы использовали базу данных «Российского мониторинга экономического положения и здоровья населения НИУ ВШЭ» (RLMS-HSE) за 2022 г. Это уникальная репрезентативная база микроданных, позволяющая учесть набор факторов, влияющих на уровень заработной платы, в этом ее основное преимущество перед агрегированными статистическими данными. При формировании выборки из общего количества респондентов были выделены занятые индивиды в трудоспособном возрасте: мужчины от 18 до 65 лет и женщины от 18 до 60 лет, получающие ненулевой трудовой доход. Также для определения инструментальных переменных была задействована база данных Росстат.

В качестве зависимой переменной мы использовали данные по заработной плате после вычета налогов. Как уже упоминалось, база РМЭЗ НИУ ВШЭ известна наличием в себе огромного числа различных переменных, и чтобы составить как можно более содержательную регрессионную модель, мы обратились к авторам, использовавшим в своих исследованиях данные РМЭЗ НИУ ВШЭ. Одним из таких авторов стала О.Н. Кадрева, которая в своей работе

[2] использует следующие независимые переменные: *количество детей меньше 18 лет, уровень образования, квадрат возраста, семейное положение*. В результате исследования подтвердилась гипотеза об отрицательном влиянии количества детей на заработную плату женщин, также, как и в работе [3] Jane Waldfogel.

Для того, чтобы построить более содержательную модель, мы добавили следующие переменные: *состояние здоровья респондента; занятость индивида в определенной отрасли экономики; тип населенного пункта, в котором проживает респондент*. Данные показатели использовал В.Н. Рудаков в своей работе «Динамика заработной платы молодежи на российском рынке труда» [5]. Для полноты анализа, мы также рассматриваем показатель *безработица в регионе*. Автор пришел к следующим выводам: как для молодежи, так и для взрослых работников высшее образование приносит существенную экономическую отдачу; влияние здоровья на заработную плату молодежи и взрослых работников оказалось незначимым. Будет интересно проверить совпадут ли результаты нашего исследования с данными выводами.

Итоговый набор переменных:

Зависимая переменная: *Wage* - *Сколько денег в течение последних 30 дней Вы получили по основному месту работы после вычета налогов?*

Независимые переменные:

1. *Age* - возраст респондентов;
2. *Age2* - квадрат возраста респондентов;
3. *Educ* - уровень образования респондентов;
4. *Status* - тип населенного пункта;
5. *Sex* - пол респондента;
6. *Marst* - семейное положение респондентов;
7. *Children* - количество детей моложе 18 лет;
8. *Health* - состояние здоровья респондентов;
9. *Height* - рост респондентов;
10. *Work_hours* - средняя продолжительность рабочей недели;

11. *Industry* - отрасль работы респондента.

12. *Unemployment_rate* - уровень безработицы по регионам (по методологии МОТ, Росстат).

Проблема мультиколлинеарности

При исследовании данных на мультиколлинеарность, не было выявлено сильно коррелирующих друг с другом признаков. Наибольшая взаимосвязь была обнаружена между признаками *Age* и *Children*: коэффициент корреляции Пирсона для этих переменных составляет -0.6 . Данная корреляция может объясняться тем, что признак *Children* отвечает за количество несовершеннолетних детей в семье, и соответственно, чем респондент старше, тем меньше у него несовершеннолетних детей. Мы не считаем, что данные признаки приведут к мультиколлинеарности, требующей борьбы с ней.

Проблема эндогенности

Проявление эндогенности - проблема гораздо серьезнее, чем проявление мультиколлинеарности. Мы предполагаем возможное наличие эндогенности среди наших данных, так как причиной могут послужить пропущенные переменные в модели или одновременность. Эндогенной может быть переменная *Health*, так как значение данного показателя определяется большим количеством факторов, и следовательно степень ее влияния на зависимую переменную может быть неверно определена. Эндогенной возможно будет являться переменная *Children*, так как наблюдается одновременность между регрессором и зависимой переменной: наличие и количество детей влияет на заработную плату [2], и в то же время размер заработной платы может влиять на количество детей в семье. Решить проблему эндогенности можно, подобрав инструментальные переменные к признакам *Health* и *Children*. Для признака *Health* мы предлагаем инструмент *ожидаемая продолжительность жизни в регионе* (*Life_exp*), так как напрямую этот показатель на заработную плату не влияет, а вот на состояние здоровья вполне (подразумеваем выполнение

релевантности и валидности инструмента). Данные были взяты с Росстата за 2022 год. Также в качестве инструмента мы используем показатель - *курит ли респондент (Smoking)*, так как курение отрицательно сказывается на здоровье человека и продолжительность жизни, следовательно, отрицательно сказывается на его способности работать по сравнению со здоровым человеком. Инструментом к переменной *Children* может быть показатель - средний размер ежемесячных выплат на ребенка по регионам (*Social_benefits*), так как при высоких значениях таких социальных выплат, люди смогут рассчитывать на поддержку со стороны государства и смогут завести еще одного или несколько детей. Для данной переменной были использованы данные членских организаций ФНПР. Также эндогенным может быть показатель *Educ* (*уровень образования*), для него мы выбрали инструмент *Education_level* - *доля занятого населения, имеющего высшее образование по регионам в 2022*. Мы предполагаем, что уровень образования респондентов зависит в том числе и от доступности образования в регионе в целом.

Инструментальные переменные:

1. *Life_exp* - ожидаемая продолжительность жизни в регионе (Росстат);
2. *Smoking* - курильщик ли респондент;
3. *Social_benefits* - ежемесячные детские выплаты в регионе (данные членских организаций ФНПР);
4. *Education_level* - доля занятого населения, имеющего высшее образование по регионам в 2022 (Росстат) ;

Предварительный анализ данных

Для улучшения качества данных, избавления от заведомо нереалистичных или нестандартных значений предложения труда (изначально занятость респондентов варьировала от 0 до 160 часов в неделю) мы удалили по 5% наблюдений на концах распределения по показателю часов работы в неделю [5]. В результате, из выборки были удалены индивиды, работающие менее 28 и

более 72 часов в неделю, а также респонденты нетрудоспособного возраста (моложе 18 или старше 65 лет). Также избавились от выбросов по 5% наблюдений на концах распределения по показателю ежемесячный доход. Теперь выборка состоит из 5251 человек. Все переменные были извлечены из данных индивидуальных опросов индивидов.

В рассматриваемой выборке средний возраст респондента составляет примерно 41 год, а продолжительность средней рабочей недели около 42 часов. Ежемесячный доход индивидов в среднем составляет 38 514 руб. в месяц (Таблица 1). Доля респондентов, проживающих в городе или областном центре превышает 70%, а в ПГТ или селе составляет около 30%. Около трети респондентов окончили среднее специальное образование, 27% имеют высшее образование, 13% - незаконченное среднее (см. рис. 1).

Рисунок 1. Распределение индивидов по уровню образования

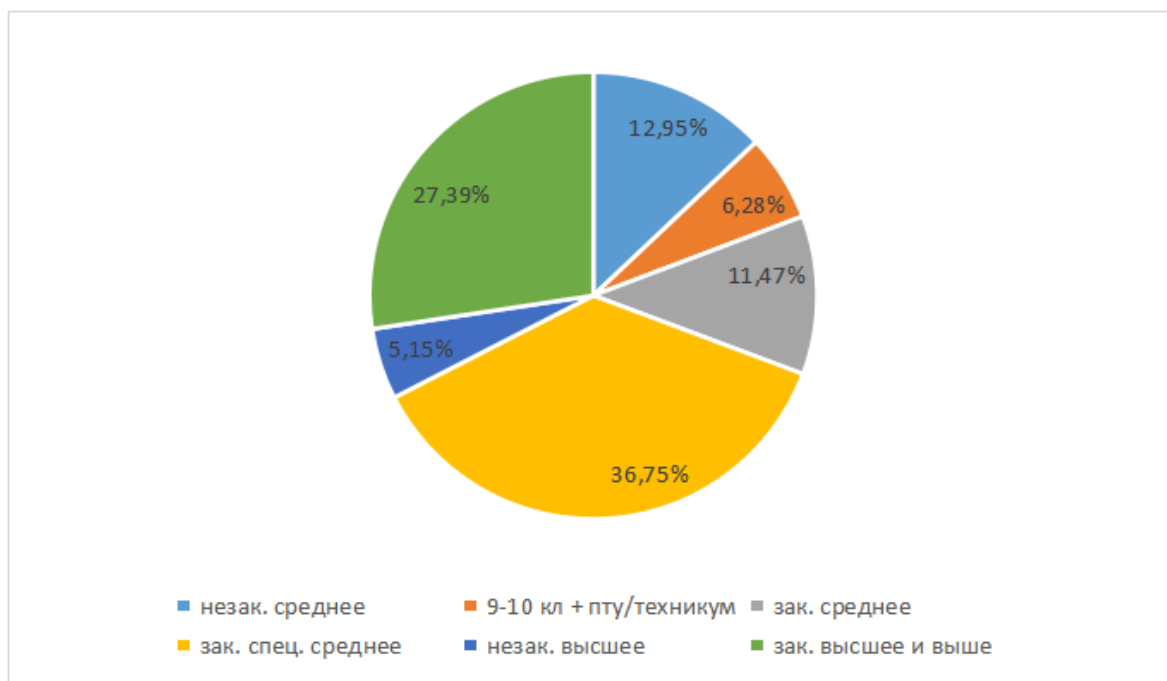


Таблица 1. Количественные объясняющие переменные и их дескриптивные статистики.

Переменная	Описание	Mean	Std. Dev.	Min	Max
Age	Возраст респондента	41.21	10.52	18	65
Children	Количество детей до 18 лет	0.98	0.97	0	9
Wage	Заработная плата после налогов	38286	15617	13000	80000
Height	Рост респондента	170.24	8.96	142	200
Work_hours	Средняя продолжительность рабочей недели	43.26	7.8	28	72
Life_exp	Ожидаемая продолжительность жизни	73.38	2.56	68.25	79.38
Education_level	Доля занятого населения, имеющего высшее образования	33.14	7.78	23.2	54.9
Unemployment rate	Уровень безработицы по регионам	3.59	1.37	1.8	10

Social_benefits	Социальные детские выплаты	8813.767	1258.979	7639.5	11754
-----------------	----------------------------------	----------	----------	--------	-------

Таблица 2. Используемые категориальные переменные.

Переменная	Описание	Значения
Health	Уровень оценки респондентом своего здоровья	1 - Очень хорошее; 2 - Хорошее; 3 - Среднее; 4 - Плохое; 5 - Совсем плохое
Status	Тип населенного пункта	1 - Областной центр; 2 - Город; 3 - ПГТ; 4 - Село
Educ	Достигнутый респондентом уровень образования	1 – незаконченное среднее; 2 – 9-10 классов + пту/техникум; 3 – законченное среднее; 4 – законченное среднее специальное; 5 – незаконченное высшее; 6 – законченное высшее и выше
Sex	Пол респондента	1 - мужской; 2 - женский
Marst	Семейное положение	1 - Никогда в браке не состояли; 2 - Состоит в зарегистрированном браке; 3 - Живете вместе, но не зарегистрированы; 4 - Разведены и в браке не состоите;

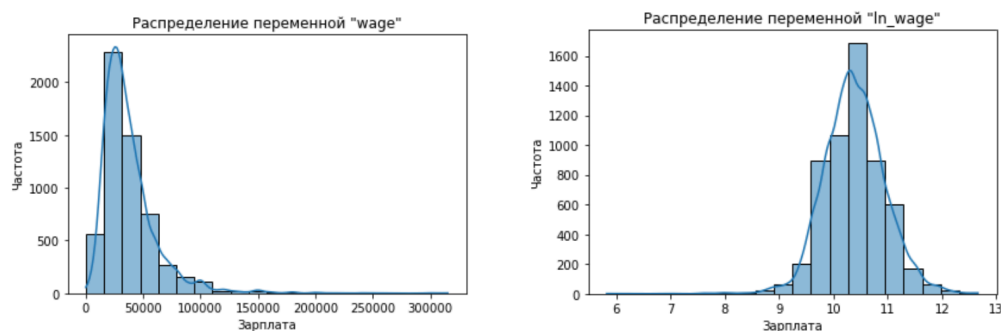
		5 - Вдовец (вдова)
Smoking	Курит ли респондент или нет	1 - Да 2 - Нет

Модель

Для выявления значимости разных факторов, которые формируют заработную плату работников на российском рынке труда, в работе будет использована модель множественной линейной регрессии: проводится оценка методом наименьших квадратов (МНК). Модель оценивается в два этапа: на первом шаге оценивается регрессия эндогенных факторов *Health*, *Educ*, *Children* на инструменты *Life_exp*, *Smoking*, *Education_level*, *Social_payments* и на экзогенные переменные. Регрессии первого шага адекватны на 5% уровне значимости. На втором шаге оценивается множественная линейная регрессия на экзогенные переменные и на переоцененные в первом шаге эндогенные переменные. Всего было построено три модели для проверки каждой из выдвинутых гипотез: общая регрессия (*гипотеза 1*); регрессия для женского пола (*гипотеза 2*); регрессия для мужского пола (*гипотеза 3*). Для сравнения моделей, оцененной с помощью МНК и двухшагового МНК с инструментальными переменными, мы провели тест Хаусмана. Результат показал, что статистически значимых различий нет, однако оцененные методом МНК оценки оказались эффективнее. Следовательно для интерпретации результатов можно использовать оценки переменных, полученных с помощью использования МНК (результаты в *Таблице 3*).

Функциональная форма: зависимая переменная в логарифмической форме. По гистограммам исходных значений и их логарифмов можно увидеть, что распределение заработных плат визуально очень похоже на логнормальное (*см рис. 2*), поэтому кажется целесообразно рассматривать именно логарифм данной переменной.

Рисунок 2. Распределение переменной “wage” и её логарифма



Также в большинстве работ, на которые мы опираемся в нашем исследовании, авторы в качестве зависимой переменной использовали натуральный логарифм заработной платы. Таким образом, уравнение регрессии выглядит следующим образом:

$$\begin{aligned} \ln_wage = & \beta_0 + \beta_1 Age2 + \beta_2 Educ + \beta_3 Health + \beta_4 Work_hours + \beta_5 Sex + \\ & \beta_6 Marst + \beta_7 Status + \beta_8 Children + \beta_9 Industry + \beta_{10} Height + \\ & \beta_{11} Unemployment_rate + \varepsilon. \end{aligned}$$

Учитывая тот факт, что исследования [2], [5] проводились на базе данных РМЭЗ НИУ ВШЭ, можно предположить, что мы получим схожие результаты: выдвинутая нами *Гипотеза 1* подтвердится частично: образование будет оказывать значимое влияние на заработную плату, а состояние здоровья нет; *Гипотеза 2* скорее всего подтвердится, так как мы используем очень схожий с О.Н. Кадревой набор независимых переменных. Результаты регрессионного анализа представлены в *Таблице 3*.

Таблица 3. Результаты регрессионного анализа

Variable	OLS_model (1)	OLS_female_model (2)	OLS_male_model (3)
health			
хорошее	-0.021	0.086	-0.072
среднее	-0.080	0.021	-0.130*

плохое	-0.218***	-0.102	-0.321***
совсем плохое	-0.358	-0.293	
status			
Город	-0.146***	-0.151***	-0.131***
ПГТ	-0.084***	-0.046	-0.124***
Село	-0.221***	-0.190***	-0.244***
marst			
Состоите в зарегистрированном браке	-0.067*	-0.116***	0.129
Живете вместе	-0.076*	-0.095**	0.093
Разведены и в браке не состоите	-0.040	-0.043	-0.026
Вдовец (вдова)	-0.015	-0.067	0.161
Официально зарег., но вместе не проживаете	-0.040	-0.063	0.066
educ			
9-10 кл. + техникум	0.052	0.013	0.077
зак. среднее	0.040	0.048	0.051
зак. среднее спец.	0.087***	0.089**	0.093***
незак. высшее	0.171***	0.213***	0.136**
зак. высшее и выше	0.292***	0.319***	0.263***
age2	-0.000**	0.000	-0.000***
sex			
женский	-0.189***		
unemployment_rate	-0.038***	-0.033***	-0.046***
work_hours	0.003***	0.002	0.005***
children	0.022***	0.010	0.036***
height	0.004***	0.004***	0.003**
industry			
гражданское машиностроение	0.091*	0.042	0.119**
ВПК	0.147***	0.177***	0.124**
др. отрасль тяжелой промышленности	0.084**	0.013	0.130***
строительство	0.070**	0.089	0.081*
транспорт, связь	0.048	-0.062	0.105***
органы управления	-0.167***	-0.236***	-0.059
образование	-0.106***	-0.130***	-0.179***
наука, культура	-0.070	-0.108*	-0.038
торговля	-0.022	-0.066*	0.056
энергетич. промышленность	0.098**	0.091	0.113*
операции с недвижимостью	0.180*	0.126	0.253

хим. промышленность	0.320***	0.366***	0.293***
услуги населению	0.054	0.183*	-0.025
IT	0.142	-0.070	0.301**
экология	-0.186	-0.366*	0.062
реклама	0.269	0.437*	-0.227
_cons	9.940***	9.667***	9.905***

legend: * p<.1; ** p<.05; *** p<.01

Интерпретация результатов

Анализируя полученные результаты, мы приходим к выводу, что *Гипотеза 1* подтверждается. Общая модель множественной регрессии показывает (*модель 1*), что качество человеческого капитала является значимым фактором на 5% уровне значимости и положительно влияет на заработную плату как мужчин, так и женщин. В понятие человеческого капитала мы включаем уровень образования и состояние здоровья респондента. Респондент, который оценивает уровень своего здоровья, как “плохое” получает на 20% меньше респондента, оценивающего здоровье, как “очень хорошее”. Однако, если посмотреть на результаты оценки регрессии отдельно для женщин (*модель 2*), то состояние здоровья оказалось незначимым фактором. Возможно такой результат связан с тем, что мужчины чаще выполняют работу, связанную с тяжелым физическим трудом, и хорошее состояние здоровья в таких профессиях напрямую влияет на размер заработной платы. Что касается уровня образования, то законченное среднее специальное позволяет повысить уровень получаемой заработной платы практически на 10%, незаконченное высшее - на 17%, а законченное высшее на целых 30%. Мы получили результаты, частично схожие с выводами в статье В.Н. Рудакова [5]: уровень образования действительно оказывает значимое влияние на заработную плату, однако, в отличие от работы [5], в нашем исследовании состояние здоровья также оказалось значимым фактором.

Интересно, что *Гипотеза 2* не подтвердилась, несмотря на то, что мы использовали схожий с О.Н. Кадревой [2] набор переменных. Исходя из модели

множественной регрессии, построенной для женщин (*модель 2*), количество детей не оказывает значимого влияния уровень заработной платы работающих женщин. Также на наших данных подтверждается факт наличия на рынке неравной оплаты труда: заработная плата мужчин выше заработной платы женщин практически на 19%. Данная закономерность не обязательно объясняется дискриминацией женского пола, можно предположить, что мужчины чаще выбирают опасные для здоровья и жизни профессии, которые выше оплачиваются. В подтверждение этой теории можно привести тот факт, что наличие высшего образования повышает доход женщин на 31.9% (*модель 2*), а доход мужчин на 26.3% (*модель 3*). То есть высшее образование для женщин важнее, чем для мужчин: высокооплачиваемые профессии, сопряженные с тяжёлым физическим трудом, не всегда требуют высокой квалификации.

Результаты модели множественной регрессии, построенной для мужчин (*модель 3*), подтвердили *Гипотезу 3*: предположение о том, что рост мужчины действительно положительно влияет на его уровень заработной платы выполняется для российского рынка труда. Исходя из результатов, представленных в *Таблице 3*, видно, что переменная *height* является значимой на 5% уровне значимости, однако влияние каждого дополнительного сантиметра незначительно и составляет всего 0.3%. Однако, если рассматривать разницу в росте в 10-20 см, то влияние становится достаточно ощутимым. Интересно, что фактор роста оказался значимым и для женщин: влияние дополнительного сантиметра в росте женщины на ее заработную плату составляет 0.4%, т.е. данный показатель влияет на доход женщин даже больше, чем на доход мужчин.

Дополнительные результаты регрессионного анализа:

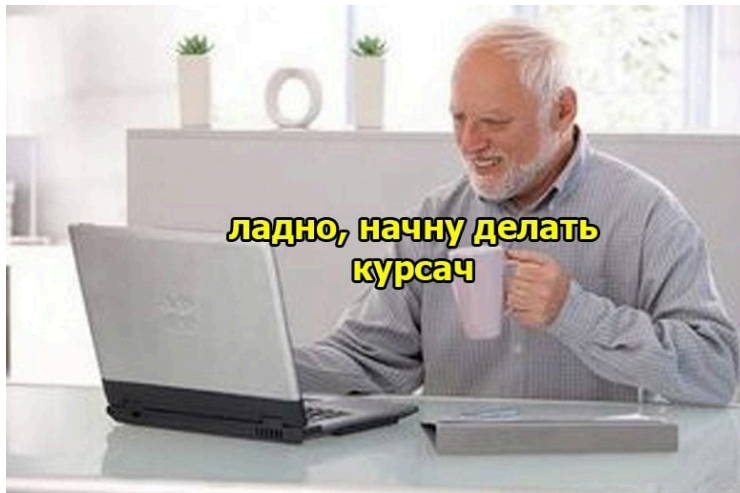
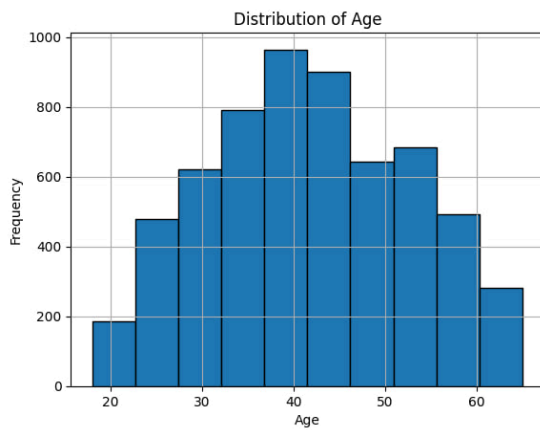
1. Респонденты, проживающие в городе, пгт или селе, получают соответственно на 15%, 10% и 20% меньше, чем респонденты, которые проживают в областном центре.

2. Респонденты, состоящие в браке или живущие вместе, но официально не зарегистрировали свои отношения, получают заработную плату примерно на 7% ниже, чем респонденты, которые никогда не состояли в браке.
3. По сравнению с оплатой труда в легкой промышленности: заработные платы в военно-промышленном комплексе выше на 14.7%; в сфере образования ниже на 10.7%; в химической промышленности выше на 32%; в органах управления ниже на 16.7%.

Список литературы

1. Васильева Е. В., Тырсин А. Н. Возрастная дифференциация в заработной плате на российском рынке труда: выявление причин //Пространственная экономика. – 2021. – Т. 17. – №. 3. – С. 103-132.
2. Кадрева О. Н. Влияние количества и возраста детей на заработки работающих женщин //Прикладная эконометрия. – 2016. – №. 1 (41). – С. 62-77.
3. Waldfogel J. The effect of children on women's wages //American sociological review. – 1997. – С. 209-217.
4. Tyrrell J. et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank //bmj. – 2016. – Т. 352.
5. Рудаков В. Н. Динамика заработной платы молодежи на российском рынке труда //Вопросы статистики. – 2015. – №. 11. – С. 43-57.

Приложение



через 10 минут:

