Федеральное бюджетное государственное образовательное учреждение высшего профессионального образования

«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА

и ГОСУДАРСТВЕННОЙ СЛУЖБЫ

при ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

**ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ**

**ОТДЕЛЕНИЕ ЭКОНОМИКИ**

Кафедра эконометрики

**Проект по эконометрике:**

**Исследование разрыва заработных плат между мужчинами и женщинами в Челябинской области.**

Выполнила:

студентка группы СП-20

Каменская Елизавета

Преподаватель:

Зямалов В.Е.

**МОСКВА, 2022 г.**

**Введение и использованные данные.**

В рамках курса эконометрики нам было предложено задание исследования разрыва между зарплатами у мужчин и женщин. Объектом этого исследования была выбрана Челябинская область. Анализ будет проводиться с помощью построения линейной регрессии с соответствующими переменными. За основу мы взяли уравнение Минцера, которое далее будем преобразовывать и проверять на разных переменных. Мы использовали данные 2020-го года, предложенные национальным исследовательским университетом ВШЭ. Для более точных результатов исключим выбросы следующим образом: нашли медиану заработных плат, вычислим межквартильный диапазон и установим границы. В следующих таблицах приведен небольшой отфильтрованный анализ населения:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Отрасль** | **Мужчин (доля населения %)** | **Женщин**  **(доля населения %)** | **Средняя ежемесячная з/п мужчин (руб/мес)** | **Средняя ежемесячная з/п женщин (руб/мес)** | **Отношение з/п женщин к з/п мужчин, %** |
| **1** | **Тяжелая промышленность** | **7(6)** | **3(2.6)** | **34285.7** | **25000** | **27** |
| **2** | **Транспорт, связь** | **11(9.4)** | **3(2.6)** | **36545.45** | **25000** | **31.6** |
| **3** | **Торговля, бытовое обслуживание** | **9(7.7)** | **18(15.4)** | **28555.5** | **22083.33** | **22.7** |
| **4** | **Образование** | **2(1.7)** | **10(8.5)** | **22000** | **21600** | **1.8** |
| **5** | **Легкая, пищевая промышленность** | **4(3.4)** | **3(2.6)** | **40750** | **32333.3** | **20.65** |
| **6** | **Юриспруденция** | **1(0.85)** | **1(0.85)** | **70000** | **70000** | **0** |
| **7** | **Строительство** | **7(6)** | **4(3.4)** | **49285.71** | **29500** | **40.14** |
| **8** | **Жилищно-коммунальное хоз-во** | **2(6)** | **2(6)** | **21500** | **34000** | **-58.14** |
| **9** | **Гражданское машиностроение** | **3(2.6)** | **1(0.85)** | **35333** | **50000** | **-41.5** |
| **10** | **Финансы** | **1(0.85)** | **1(0.85)** | **25000** | **35000** | **-40** |
| **11** | **Энергетичсекая промышленность** | **2(6)** | **2(6)** | **65000** | **28500** | **56** |
| **12** | Армия, МВД | **3(2.6)** | **2(6)** | **19333.3** | **32500** | **-68.1** |
| **13** | Операции с недвижимостью | **0(0)** | **1(0.85)** |  | **30000** |  |
| **14** | Здравоохранение | **0(0)** | **9(7.7)** | **31777** |  |  |
| **15** | Нефть и газ | **1(0.85)** | **0(0)** | **35000** |  |  |
| **16** | Военная промышленность | **0(0)** | **1(0.85)** |  | **34000** |  |
| **17** | Наука, культура | **0(0)** | **2(6)** |  | **17500** |  |
| **18** | IT | **1(0.85)** | **0(0)** | **30000** |  |  |

**Всего в опросе участвовало 117 человек. Из них: 51 мужчина и 66 женщин.**

Из этой таблицы видно, что самые высокие зарплаты и мужчины и женщины получают, работая в юриспруденции, самые низкие - в отрасли образования. В строительстве мужчины на 40% получают зп больше, чем женщины, женщины на 68.1% получают зп больше чем мужчины, работая в армии, так же на 58% они получают больше чем мужчины, работая в сфере ЖКХ. Также интересно отметить, что в сфере образования, здравоохранения и бытовых услуг женщин работает гораздо больше чем мужчин, чего не скажешь о сферах тяжелой промышленности и транспорта.

Данных достаточно мало и не для всех сфер деятельности нашлись респонденты, которые участвовали в опросе или их очень мало, это может привести к неточным результатам.

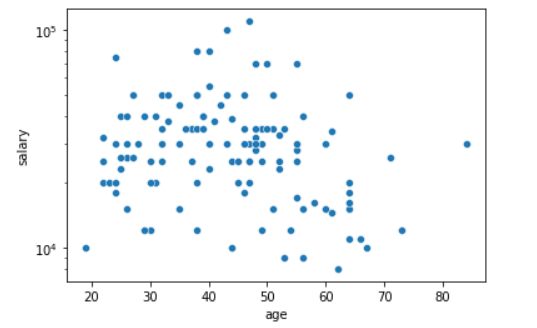
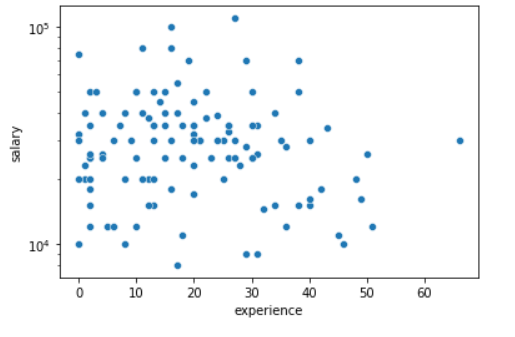
**Первая построенная модель:**

Как уже было сказано, основой модели стало уравнение Минцера. Оно имеет следующий вид:

Здесь

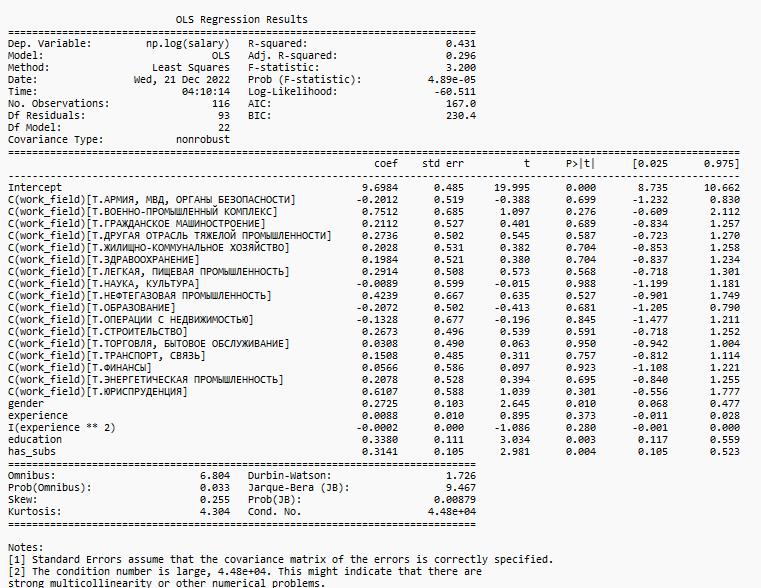
* lnw – натуральный логарифм среднемесячной заработной платы;
* age – возраст работника;
* exp – стаж работы;
* gend – пол работника (1 – мужчина; 0 – женщина) ;
* stat– тип населенного пункта (1 – город; 0 – село)edu – уровень образования:
* edu - (1 – высшее; 0 – остальные);

Протестируем еще две альтернативные модели: модель с учетом возраста и модель с учетом к-ва детей у женщин.

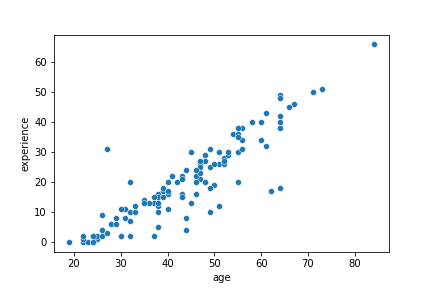


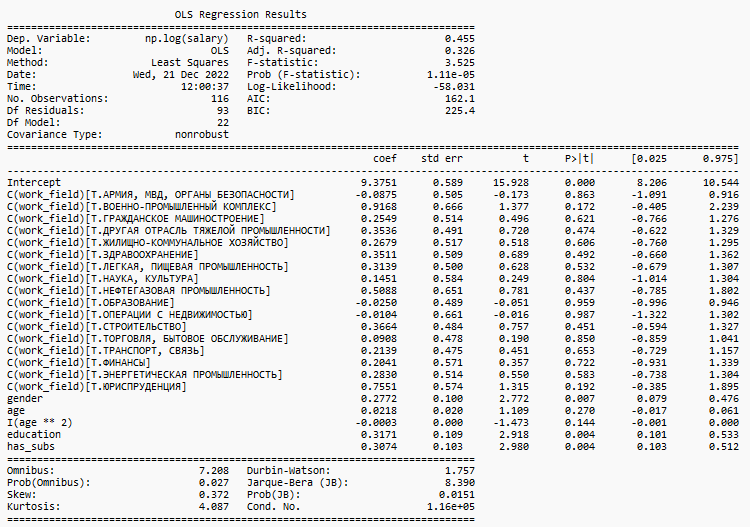
На диаграммах рассеяния видно, что у первой модели экспериментальные точки больше упорядочены как линейная зависимость, поэтому предполагаем, что перваяя модель окажется лучше.

МНК-оценки:



Есть все основания полагать, что возраст и опыт нельзя использовать как объясняющие переменные в одном уравнении. Очевидно, они имеют линейную зависимость:

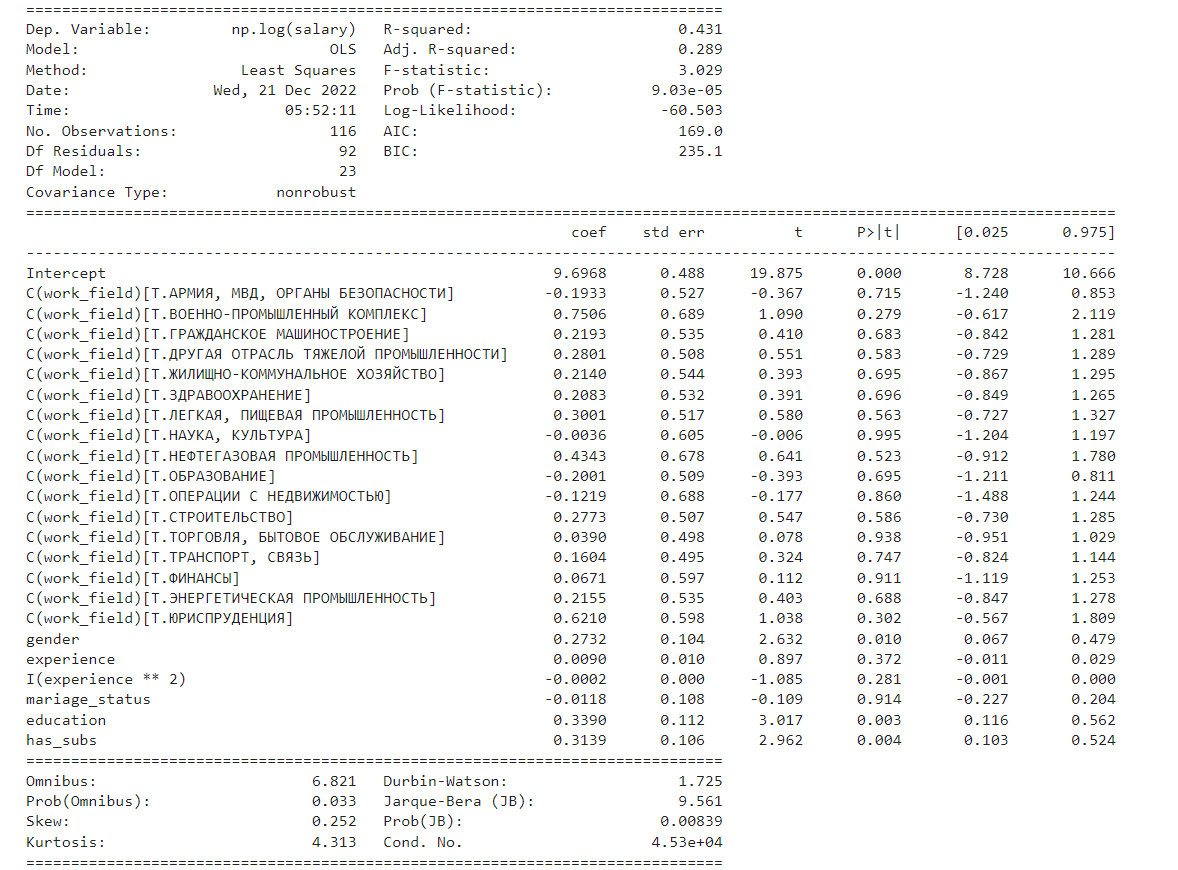


Следующие оценки приведены для модели, где вместо опыта мы включаем возраст агента:

Вторая модель учитывает семейное положение у женщин. Таким образом она принимает следующий вид:

Здесь появилась еще одна переменная – get\_married

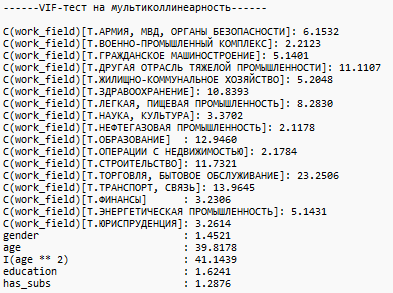
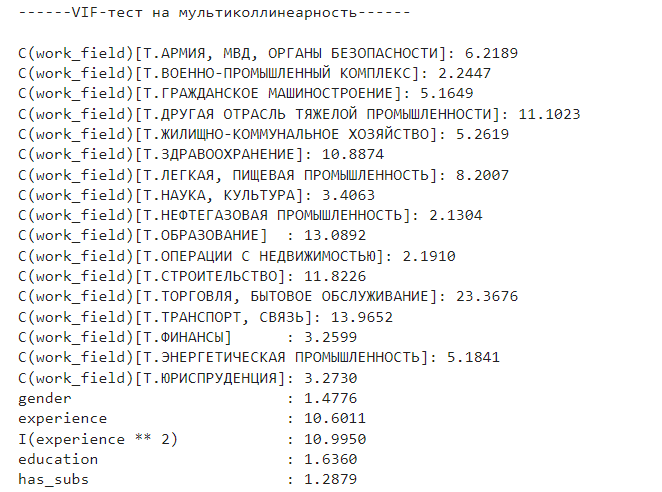
Оценки для этой модели:



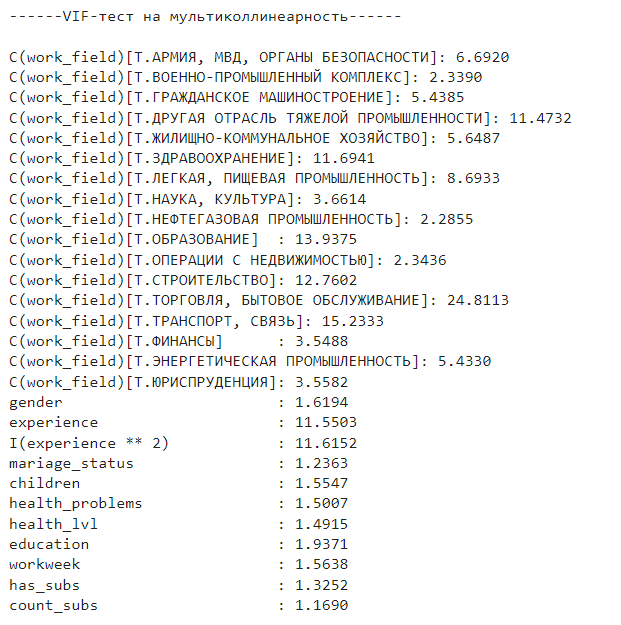
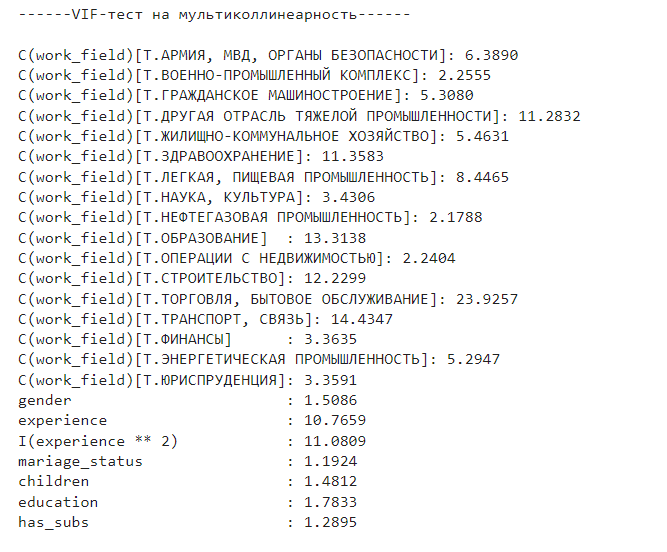
Видим, что к-во детей все же почти не повлияло на результаты, хотя это кажется достаточно весомым аргументом. Также на результаты почти не влияет добавление и остальных переменных(таких как уровень здоровья, семейное положение и проч.), что говорит о достаточности построенной модели. При переменной gender получаем оценку коэффициента на уровне 0.26. Это говорит о том, что средняя зарплата женщин меньше на 26% при прочих равных. Этот коэффициент значим на 1% уровне значимости.

**Проверка на мультиколлинеарность:**

Наличие мультиколлинеарности мы проверяли при помощи vif-теста

Для первой и второй модели: 

Для третьей модели и модели со всеми возможными переменными:



Как видим, каждый тест дает результаты, говорящие о наличии мультиколлинеарности. Однако, эти результаты связаны с присутствующими функциями от опыта или возраста, что обязательно должно было присутствовать. Поэтому, в целом, будем считать что мультиколлинеарности нет.

**Проверка условий Гаусса-Маркова**

Проверка гетероскедастичности:

Ошибки на гомоскедастичность проверены с помощью теста Бройша-Пагана.

**Нулевая гипотеза (H0 ):** присутствует гомоскедастичность.

**Альтернативная гипотеза: (Ha):** гомоскедастичность *отсутствует* (т.е. гетероскедастичность существует)

Результат теста (значения f p-value):

1)0.833

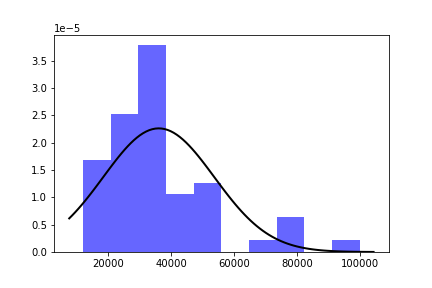
2) 0.822

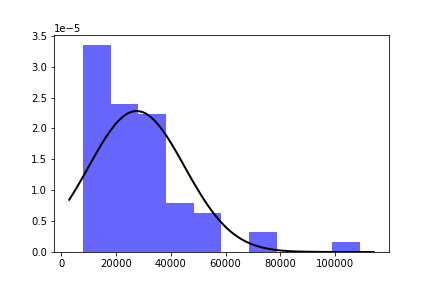
3)0.853744

Нулевая гипотеза не отвергнута, т.к. все значения не меньше 0.05

**Проверка на распределение:**

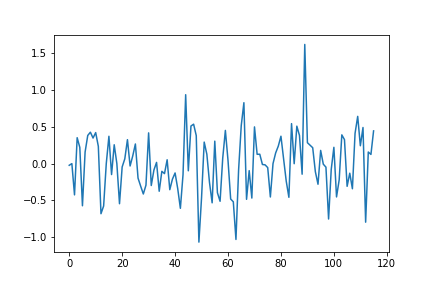
Распределение зарплат по мужчинам:



и распределение зарплат по женщинам:

Как видим, распределение зарплат у мужчин больше похоже на нормальное, чем у женщин. Это объясняется тем, что в нашей выборке мужчины затрагивают больше профессий. Большее к-во данных дало бы лучшие результаты.

**Проверка матожидания ошибок**

Построим динамику ошибок по имеющимся точкам. Видим, что они колеблются около нуля, поэтому считаем, что матожидание ошибки примерно ноль.

**Выводы**

Были построены три модели для изучения разрыва заработных плат между мужчинами и женщинами. Полученные результаты говорят о разрыве в 27% (в среднем по всем моделям)), при этом мужчины получают больше зарплату, что является значимым результатом. Наилучшей моделью оказалась первая, т.к. она дает наилучшие результаты по vif-тесту и оценки для нее могут быть отвергнуты на меньших уровнях значимости.