­­­ Федеральное государственное образовательное бюджетное учреждение высшего образования

**«Финансовый университет при Правительстве Российской Федерации»**

**(Финансовый университет)**

Факультет «Информационных технологий и анализа больших данных»

Направление подготовки

01.03.02 Прикладная математика и информатика

Анализ данных и принятие решений в экономике и финансах

***Индивидуальное домашнее задание***

По дисциплине «Эконометрика» на тему:

«Анализ факторов, влияющих на зарплату»

Выполнил студент

3 курса, учебной группы ПМ21-2

Меньшиков Илья Владимирович

Москва 2023

Анализ факторов, влияющих на зарплату

**1. Постановка задачи**

Проанализировать факторы влияющие на зарплаты с сайта hh.ru

**2. Описание используемых данных**

**2.1. Источник данных**

Поиск готовых, качественных и актуальных данных оказался непростым. Данные были получены с сайта hh.ru. Для этого был использован проект, найденный на github.com. Этот проект позволяет указать нужные параметры и сохранить данные в формате csv.

**2.2. Описание данных**

Было решено использовать вакансии на позицию 1С программиста. После парсинга было получено 851 вакансий.

Датасет содержит факторы:

* **id** - идентификатор
* **published\_at** - дата публикации
* **name** - название вакансии
* **area** - город
* **employer** - работодатель
* **type** - тип вакансии
* **experience** - опыт работы
* **description** - описание вакансии
* **key\_skills** - ключевые навыки
* **alternate\_url** - ссылка
* **schedule** - график работы
* **employment** - вид занятости
* **contacts** - контактная информация
* **description\_cleaned** - очищенное описание
* **salary\_from** - зарплата от
* **salary\_to** - зарплата до

Часть из них являются ненужными для дальнейшей работы, поэтому их следует исключить, а некоторые преобразовать.

Зависимая переменная: **Y** – зарплата.

После преобразований остались следующие факторы:

**X1(Experience)** – 0 (нет опыта), 1 (опыт от 1 до 3 лет), 2 (опыт от 3 до 6 лет), 3 (опыт более 6 лет)

**X2(is\_Moscow)** - бинарная переменная, где за 1 я принял г. Москва, а за 0 остальные города.

**X3(is\_StPetersburg)** – бинарная переменную, где за 1 я принял г. Санкт-Петербург, а за 0 остальные города.

**X4(schedule\_Полный день)** – бинарная переменная, где за 1 я принял полный рабочий день, а за 0 удаленную работу

**3. Анализ факторов указанных в вакансиях**

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

Общее распределение вакансий:

* Без опыта: 14.1%
* 1-3 года опыта: 36%
* 3-6 лет опыта: 43.8%
* Более 6 лет опыта: 6.1%

Тенденции:

* Спрос растет с увеличением опыта.
* Высокий спрос на опыт 3-6 лет (43.8%).
* Низкий спрос на более опытных (6.1%).

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

* Москва лидирует по количеству вакансий, предоставляет гораздо больше возможностей.
* Санкт-Петербург имеет ограниченные варианты для категорий без опыта и с опытом более 6 лет.
* Для кандидатов с 6+ лет опыта, а также без опыта стоит рассматривать только Москву и Санкт-Петербург, в других городах почти нет вакансий.

Изображение выглядит как текст, снимок экрана, диаграмма, Шрифт

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, диаграмма, Прямоугольник

Автоматически созданное описание

* Для тех, кто ищет полный рабочий день, наибольшее количество вакансий доступно, особенно для категорий 3-6 лет опыта и 1-3 года.
* Удаленная работа предоставляет некоторые возможности, особенно для кандидатов с опытом 3-6 лет, но варианты для без опыта и с 6+ лет ограничены.
* Гибкий график предоставляет минимальное количество вакансий, и варианты для кандидатов с 6+ лет опыта отсутствуют.

Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описание

Компания «Первый Бит» ищет сотрудников на все уровни опыта: без опыта, с опытом 1-3 лет и с опытом 3-6 лет.

Компания «Крон» фокусируется на поиске сотрудников только с опытом 3-6 лет.

Компания «1С» также ищет сотрудников на различные уровни опыта: с опытом 1-3 лет, 3-6 лет, а также с более чем 6 лет опыта, не рассматривая кандидатов без опыта.

Изображение выглядит как текст, снимок экрана, Шрифт, диаграмма

Автоматически созданное описание

Большинство компаний 59.8%, предоставляющих вакансии, указывают зарплатные условия, однако 40.2% вакансий не предоставляют информацию о зарплате.

Изображение выглядит как диаграмма, снимок экрана, линия, прямоугольный

Автоматически созданное описание

**Без опыта:**

* Средняя зарплата составляет примерно 60,789 рублей.
* Медианное значение равно 46,650 рублей.
* Разброс зарплат от 15,000 до 220,000 рублей.
* Основная масса значений сконцентрирована в нижней половине диапазона, с несколькими выбросами вверх.

**Опыт 1-3 года:**

* Средняя зарплата значительно выше: примерно 137,098 рублей.
* Медианное значение составляет 130,000 рублей.
* Разброс зарплат от 27,000 до 300,000 рублей.

**Опыт 3-6 лет:**

* Средняя зарплата продолжает увеличиваться и составляет около 192,200 рублей.
* Медианное значение равно 180,000 рублей.
* Разброс зарплат от 40,000 до 500,000 рублей.

**Опыт 6+ лет:**

* Средняя зарплата дальше увеличивается и составляет примерно 199,444 рубля.
* Медианное значение равно 210,000 рублей.
* Разброс зарплат от 10,000 до 350,000 рублей.

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, линия, число

Автоматически созданное описание

**4. Расчет параметров регрессионного уравнения**

Была построена матрица парных коэффициентов корреляции для оценки значимости этих факторов.

Изображение выглядит как текст, снимок экрана

Автоматически созданное описание

Красным цветом выделена сильная корреляционную связь.

Анализ матрицы парных коэффициентов корреляции показывает, что:

* Опыт работы (experience) положительно коррелирует с нахождением в Москве и средней зарплатой.
* График работы "Полный день" слабо отрицательно коррелирует с опытом работы и не имеет явной связи с городом.
* Москва положительно коррелирует с опытом работы и средней зарплатой.
* Санкт-Петербург слабо коррелирует с опытом работы и средней зарплатой.
* Средняя зарплата сильно положительно коррелирует с опытом работы и умеренно с нахождением в Москве, слабо с нахождением в Санкт-Петербурге.

Чтобы проверить значимость коэффициентов корреляции я использую t-критерий Стьюдента. Табличное значение t-статистики Стьюдента с уровнем значимости 0,05.

Таблица 2 Регрессионный анализ

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Все факторы, кроме schedule\_Полный день признаются статистически значимыми, что свидетельствуют о наличии устойчивой связи.

После исключения schedule\_Полный день:

Изображение выглядит как текст, снимок экрана, число, Параллельный

Автоматически созданное описание

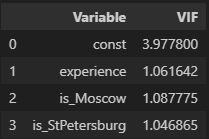
Изображение выглядит как текст, снимок экрана, меню, Шрифт

Автоматически созданное описание

Все факторы значимы.

Проверка модели на мультиколлиниарность.

Для проверки использую VIF-тест



Результаты VIF-теста показывают, что между независимыми переменными нет мультиколлинеарности. Все значения VIF находятся на уровне, менее 5.

**Ŷ = 58389.2 + 50929.66\*X1 + 68501.98\*X2 + 33856.1\* X3,**

Где X1 – experience, X2 - is\_Moscow, X3 - is\_StPetersburg

**5. Анализ адекватности регрессионного уравнения**

**5.1. Показатели качества подгонки**

Регрессионное уравнение успешно объясняет примерно 54.7% вариации в зависимой переменной (R² = 0.547). Это говорит о том, что модель в значительной степени учитывает изменения в данных.

Значение F-статистики (178.8) свидетельствуют о статистической значимости регрессии в целом. Модель имеет обоснованность в использовании для описания данных.

Все коэффициенты модели являются статистически значимыми при уровне значимости 0.05, что подчеркивает их важность в объяснении средней зарплаты.

Таким образом, данное уравнение регрессии адекватно и может быть использовано для прогнозирования зависимой переменной на основе значений факторов.

**5.2. Проверка условий для получения «хороших» оценок МНК**

H0: данные соответствуют нормальному распределению.

H1: данные не соответствуют нормальному распределению.

Тест Хархе-Бера (JB) показывает значение статистики 1.764 и p-value (Prob(JB)) 0.414. Поскольку p-value больше уровня значимости, то H0 не отвергается. Таким образом, на имеющихся данных не обнаружено статистически значимых отклонений от нормального распределения остатков.

Изображение выглядит как линия, текст, График, диаграмма

Автоматически созданное описание

Для проверки на гетероскедастичность использовал тест Голдфелда-Квандта.

H0: присутствует гомоскедастичность.

H1: присутствует гетероскедастичность.

Результат теста Голдфелда-Квандта показывает, что p-value равно 0.011, что меньше уровня значимости 0.05. Таким образом, на уровне значимости 0.05 у нас есть статистически значимые доказательства гетероскедастичности в остатках.

**5.3. Экономический смысл коэффициентов регрессии**

Коэффициенты эластичности позволяют определить силу влияния каждого фактора на цену квартиры. Полученные значения эластичности для каждого фактора составляют:

**Э1 = 0.411955** (const) представляет базовый уровень зарплаты, когда все независимые переменные равны нулю. Без опыта, не из Москвы и не из Санкт-Петербурга.

**Э2 = 0.452365** (experience) указывает, что увеличение опыта на одну категорию (отсутствие опыта, опыт 1-3 года, опыт 3-6 лет, более 6 лет) соответствует увеличению цены вакансии на 45%. Это указывает на то, что более опытные кандидаты требуют более высокую зарплату.

**Э3 = 0.114353** (is\_Moscow) указывает, что вакансии в Москве, по сравнению с другими городами, имеют ценовое преимущество на 11%.

**Э4 = 0.021327** (is\_StPetersburg) указывает, что вакансии в Санкт-Петербурге, по сравнению с другими городами, имеют ценовое преимущество на 2%.

**Корень из среднеквадратической ошибки (RMSE):** Значение RMSE равно 48022.49, средняя ошибка зарплаты в пределах 48022.49 рублей.

**Средняя абсолютная ошибка (MAE):** Значение MAE составляет 38945.72, средняя абсолютная ошибка зарплаты 38945.72.

**6. Выводы**

В результате проведенной работы были определены, а также проанализированы факторы влияющие на зарплату, построена модель для прогнозирования.

В результате анализа факторов, влияющих на зарплаты вакансий 1С программиста, было установлено, что опыт работы, местоположение (особенно Москва) и тип графика работы значительно влияют на уровень заработной платы. Построенно регрессионное уравнение, проведен анализ адекватности уравнения и проверен экономический смысл его коэффициентов.

**Используемые источники:**

1. <https://github.com/coderosemarisha/hh> Данный репозиторий предоставляет инструменты и возможности для получения данных с сайта.

2. hh.ru. Это один из ведущих онлайн-порталов по поиску работы.

3. Учебник "Эконометрика" И. И. Елисеева, 2012 г.