Федеральное государственное бюджетное образовательное учреждение высшего образования «Уфимский университет науки и технологий»

Институт экономики, управления и бизнеса

кафедра цифровой экономики и коммуникаций

**ОТЧЕТ**

**по лабораторной работе №1**

Линейная регрессия средствами Excel

**Выполнил:** Студент Ахметов Руслан группа 2ФКЦ

**Проверил:** асс. Владимирова И.П.

**Уфа 2023**

# Лабораторная работа №1. Парная регрессия

**Цель работы**

Рассчитать параметры парного линейного уравнения регрессии с помощью MS Excel, а также проанализировать качество построенной модели, используя коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

**Задание**

1. Построить линейное уравнение парной регрессии *y* по *x*.
2. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью *F*-критерия Фишера и *t*-критерия Стьюдента.
4. На одном графике отложить исходные данные и теоретическую прямую.

**Выполнение работы:**

1. Находим параметры уравнения регрессии a и b



Получено уравнение регрессии: **y= -1153,08+2,65x**

**Этапы проверки адекватности:**

***1 Этап. Проверка общего качества уравнения регрессии***

Проверка общего качества уравнения регрессии производится на основе показателей качества подгонки. Показатели качества подгонки:

остаточная дисперсия – несмещенная оценка дисперсии случайной ошибки линейной модели Sigma^2



Затем находим коэфицент корреляции r\_xy



r\_xy равняется **0.57** это показатель показывает, что связь **ЗАМЕТНАЯ**

Значения коэффициента могут варьировать в пределах [-1; 1]. Чем ближе  к 1 тем более тесная связь между изучаемыми признаками. , означает, что отсутствует линейная зависимость, однако между признаками может иметь место нелинейная зависимость.

Теперь мы найдем коэфицент детерминации R



Коэффициент детерминации (характеризует долю вариации результативного признака, обусловленную вариацией фактора): нашем примере 32%. Это означает, что расчетные параметры модели на 32% объясняют зависимость между изучаемыми параметрами. Вариация среднедневной заработной платы на 32% объясняется вариаций среднедушевого прожиточного минимума, остальное – это параметры, не включенные в модель или случайная ошибка. Так же следует отметить, что чем выше коэффициент детерминации, тем качественнее модель.

Ошибка аппроксимации



Можно сказать, что качество построенной модели не совсем хорошее, так как превышает 10%

Информационные критерии:



Информационные критерии используются для сравнения моделей между собой. Принято, что чем меньше показатель моделей, тем лучше. В нашем случае можно сказать, что они достаточно высокие.

***Этап 2. Проверка статистической значимости уравнения регрессии***

Оценку статистической значимости уравнения регрессии в целом проведем с помощью F-критерия Фишера, расчетное значение F-критерия составит **8,13**. Теперь мы рассчитаем F\_табл, оно же у нас составит **4,45**.

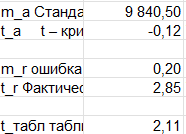
Приведем 2 гипотезы

1. H0: b=0, то есть коэф. При регрессорах равны 0 и значит уравнение не значимо. 2) H1: b ≠ 0, то есть линейная связь между переменными существует и уравнение значимо.

**Так как модуль F\_расч > F\_табл, то мы отклоняем гипотезу H0, в пользу H1.**

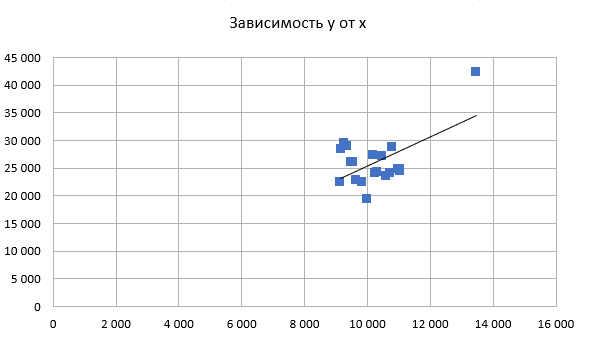
Оценим отдельные параметры уравнения регрессии:

Находим стандартную ошибку для коэффициентов a, b и корреляции r\_xy. Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, то есть определяется фактическое значение t – критерия Стьюдентакоторое затем сравнивается с табличным значением при определенном уровне значимости альфа и числе степеней свободы n-2.



Можно сказать, что коэффициент **b** и коэффициент **r\_xy** – значимы, а коэффициент **a** является не значимым, так как значение **t\_a < t\_табл.**

Построим график используя исходные данные:



***Этап 3. Проверка выполняемости условий классической линейной регрессионной модели***

Предпосылки МНК:

1.  (математическое ожидание остатков равно нулю)
2.  (дисперсия остатков постоянна)
3.  остатки не коррелированы
4.  (остатки есть нормально распределенная случайная величина);

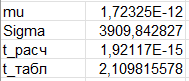
**a)** **Проверка условия** **.**





Тестовая статистика, которая подчиняется t-распределению Стъюдента. Находим все необходимые данные для расчета F\_расч, а именно  – несмещенное выборочное стандартное отклонение остатков;  – выборочное среднее остатков.

Находим F\_расч и F\_табл



По скольку F\_расч < F\_табл, то мы принимаем гипотезу H0, следовательно мат ожидание ровно 0.

**б) Проверка условия**  **(дисперсия остатков постоянна).**

Постоянство дисперсии или гомоскедастичность остатков необходимое условие при проверке достоверности построенной модели.

Гетероскедастичность – нарушение условия постоянства дисперсии случайных ошибок.

**Тест ранговой корреляции Спирмена**

Считается, что данный тест применяется при малых объемах выборки и произвольном законе распределения остатков.

Тест основан на показателе ранговой корреляции Спирмена, который рассчитывается по формуле:

,

где - абсолютная разность между рангами значений и

Если наблюдаемое значение t-статистики, вычисленное по формуле: не превышает табличное значение (определяется по таблице критических значений распределения Стьюдента на основании уровня значимости и числа степеней свободы), то гипотезу об отсутствии гетероскедастичности отклоняют. В противном случае гипотеза об отсутствии гетероскедастичности принимается. Делаем рассчеты.



Поскольку t\_расч < t\_набл, то мы принимаем гипотезу H0, следовательно D(esp) = const

**в)**  **остатки не коррелированы**

Автокорреляция остатков – нарушения условий независимости между остатками для разных наблюдений, т.е. 

**Тест Дарбина-Уотсона**

Выдвигаются гипотезы:

{Автокорреляция отсутствует}

{Наличие автокорреляции}

Далее по таблице (Значения статистик Дарбина – Уотсона *dL dU* при 5%-ном уровне значимости) определяются критические значения критерия Дарбина–Уотсона  и  (наблюдений *n*, числа независимых переменных модели *k* и уровня значимости α. По этим значениям числовой промежуток [0;4] разбивают на пять отрезков.

****

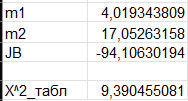
Рассчитываем d и по таблице берем dl и du



Так как **,** то принимается гипотеза об отсутствии автокорреляции.

**г)  (остатки есть нормально распределенная случайная величина).**

Если случайная компонента распределена нормально, то и коэффициенты регрессии будут распределены также. Если же условие данное не выполняется, то оценки являются не состоятельными. Для проверки данный гипотезы используется критерий Харке-Бера. В качестве основной гипотезы выдвигается гипотеза Н0 о «нормальности» распределения остатков, при этом альтернативной гипотезой является Н1 остатки не распределены по нормальному закону распределения.



Т.к**.**  то принимаем  о соответствии распределения остатков нормальному распределению на уровне значимости α=0,05.

Все предпосылки условий классической линейной регрессионной модели выполнены.

**Этап 4. Экономическая интерпретация модели**

Параметр регрессии позволяет сделать вывод, что с увеличением величины прожиточного минимума трудоспособного населения на 1 руб, среднедушевые денежные доходы населения по субъектам РФ, возрастает в среднем на **2,65** руб.