

Федеральное государственное образовательное бюджетное учреждение  
высшего образования  
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ  
РОССИЙСКОЙ ФЕДЕРАЦИИ»**  
Новороссийский филиал  
Кафедра «Информатики, математики и общегуманитарные науки»

**И.Г. РЗУН**  
**Методические рекомендации**  
**ЭКОНОМЕТРИКА**

Направление подготовки: 38.03.01 Экономика  
Направленность (профиль): Финансы и кредит  
Форма обучения: очная, заочная  
Квалификация (степень) выпускника: бакалавр

Новороссийск 2020

# **МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ИЗУЧЕНИЮ ДИСЦИПЛИНЫ**

## **Цели и задачи**

Целью изучения дисциплины является формирование у обучающихся системного представления об эконометрике, как науке, исследующей данные статистики для изучения поведения, описания и прогнозирования развития различных факторов.

## **Задачи дисциплины**

Важной методической задачей курса является формирование понимания обучающимися основных положений эконометрики; приобретение опыта построения эконометрических моделей, принятия решений о спецификации и идентификации модели и выбора метода оценки параметров модели, интерпретации результатов, получения прогнозных оценок на основе анализа эконометрических данных; освоение современных эконометрических пакетов прикладных программ.

## **РЕКОМЕНДАЦИИ ПО ПРОВЕДЕНИЮ ЛЕКЦИЙ**

При подготовке к занятиям обучающийся должен просмотреть конспекты лекций, рекомендованную литературу по данной теме; подготовиться к ответу на контрольные вопросы. Успешное изучение дисциплины требует от обучающихся посещения лекций, активной

работы на семинарах, выполнения всех учебных заданий преподавателя, ознакомления с основной и дополнительной литературой, интернет-источниками.

Запись лекции – одна из форм активной самостоятельной работы обучающихся, требующая навыков и умения кратко, схематично, последовательно и логично фиксировать основные положения, выводы, обобщения, формулировки. Культура записи лекции – один из важнейших факторов успешного и творческого овладения знаниями. Последующая работа над текстом лекции воскрешает в памяти содержание лекции, позволяет развивать аналитическое мышление. Лекции имеют обзорный характер и нацелены на освещение наиболее трудных и дискуссионных вопросов, а также призваны способствовать формированию навыков самостоятельной работы с научной литературой. Работа с конспектом лекций предполагает просмотр конспекта в тот же день после занятий, пометку материала конспекта, который вызывает затруднения для понимания. Попробуйте найти ответы на затруднительные вопросы, используя рекомендуемую литературу. Если самостоятельно не удалось разобраться в материале, сформулируйте вопросы и обратитесь за помощью к преподавателю на консультации, ближайшей лекции или семинаре. Регулярно отводите время для повторения пройденного материала, проверяя свои знания, умения и навыки по контрольным вопросам.

Работу с основной и дополнительной литературой целесообразно начинать с освоения материала учебников, которые содержат необходимый материал по каждой теме.

Подготовка к семинарскому занятию зависит от темы занятия и вопросов, предложенных преподавателем, для подготовки к семинару.

Постоянная активность на занятиях, готовность ставить и обсуждать актуальные проблемы дисциплины – залог успешной работы и положительной оценки.

## ЛИНЕЙНЫЙ ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Одним из методов изучения стохастических связей между признаками является регрессионный анализ.

Регрессионный анализ представляет собой вывод уравнения регрессии, с помощью которого находится средняя величина случайной переменной (признака-результата), если величина другой (или других) переменных (признаков-факторов) известна. Он включает следующие этапы:

- 1) выбор формы связи (вида аналитического уравнения регрессии);
- 2) оценку параметров уравнения;
- 3) оценку качества аналитического уравнения регрессии.

Наиболее часто для описания статистической связи признаков используется линейная форма. Внимание к линейной связи объясняется четкой экономической интерпретацией ее параметров, ограниченной вариацией переменных и тем, что в большинстве случаев нелинейные формы связи для выполнения расчетов преобразуют (путем

логарифмирования или замены переменных) в линейную форму.

В случае линейной парной связи уравнение регрессии примет вид:  $y_i = a + b \cdot x_i + u_i$ . Параметры данного уравнения  $a$  и  $b$  оцениваются по данным статистического наблюдения  $x$  и  $y$ . Результатом такой

оценки является уравнение:  $\tilde{y}_i = \tilde{a} + \tilde{b} \cdot x_i$ , где  $\tilde{a}$ ,  $\tilde{b}$

- оценки параметров  $a$  и  $b$ ,  $\tilde{y}_i$  - значение результирующего признака (переменной), полученное по уравнению регрессии (расчетное значение).

Наиболее часто для оценки параметров используют метод наименьших квадратов (МНК).

Метод наименьших квадратов дает наилучшие (состоятельные, эффективные и несмещенные) оценки параметров уравнения регрессии. Но только в том случае, если выполняются определенные предпосылки относительно случайного члена ( $u$ ) и независимой переменной ( $x$ ).

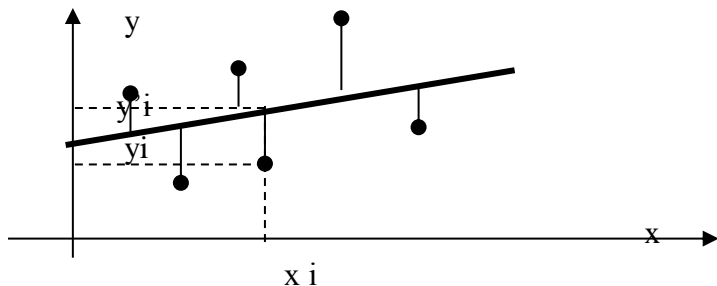
Задача оценивания параметров линейного парного уравнения методом наименьших квадратов состоит в следующем:

получить такие оценки параметров  $\tilde{a}$ ,  $\tilde{b}$ , при которых сумма квадратов отклонений фактических значений результирующего признака -  $y_i$  от расчетных значений -  $\tilde{y}_i$  минимальна.

Формально критерий МНК можно записать так:

$$S = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \rightarrow \min$$

Проиллюстрируем суть данного метода графически. Для этого построим точечный график по данным наблюдений  $(x_i, y_i, i=1;n)$  в прямоугольной системе координат (такой точечный график называют корреляционным полем). Попытаемся подобрать прямую линию, которая ближе всего расположена к точкам корреляционного поля. Согласно методу наименьших квадратов линия выбирается так, чтобы сумма квадратов расстояний по вертикали между точками корреляционного поля и этой линией была бы минимальной.



Математическая запись данной задачи:

$$S = \sum_{i=1}^n (y_i - (\tilde{a} + \tilde{b} \cdot x_i))^2 \rightarrow \min$$

Значения  $y_i$  и  $x_i, i=1;n$  нам известны, это данные наблюдений. В функции  $S$  они представляют собой константы. Переменными в данной функции являются искомые оценки параметров -  $\tilde{a}, \tilde{b}$ . Чтобы найти минимум функции 2-ух переменных необходимо вычислить частные производные данной функции по

каждому из параметров и приравнять их нулю, т.е.

$$\frac{\partial S}{\partial \tilde{a}} = 0, \quad \frac{\partial S}{\partial \tilde{b}} = 0$$

В результате получим систему из 2-ух нормальных

$$\left\{ \begin{array}{l} \sum_{i=1}^n y_i = \tilde{a} \cdot n + \tilde{b} \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i \cdot x_i = \tilde{a} \sum_{i=1}^n x_i + \tilde{b} \sum_{i=1}^n x_i^2 \end{array} \right.$$

линейных уравнений:

Решая данную систему, найдем искомые оценки параметров:

$$\tilde{b}_1 = \frac{n \sum x_i y_i - \bar{x} \cdot \bar{y}}{n \sum x_i^2 - (\bar{x})^2} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$$

$$\tilde{a} = \bar{y} - \tilde{b} \cdot \bar{x}$$

Правильность расчета параметров уравнения регрессии может быть проверена сравнением сумм

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \tilde{y}_i \quad (\text{возможно некоторое расхождение из-за округления расчетов}).$$

Для расчета оценок параметров  $\tilde{a}$ ,  $\tilde{b}$  можно построить таблицу 1.

Знак коэффициента регрессии  $\tilde{b}$  указывает направление связи (если  $\tilde{b} > 0$ , связь прямая, если  $\tilde{b} < 0$ , то связь обратная). Величина  $\tilde{b}$  показывает на сколько единиц изменится в среднем признак-результат -у при изменении признака-фактора - x на 1 единицу своего измерения.

Формально значение параметра  $a$  – среднее значение  $y$  при  $x$  равном нулю. Если признак-фактор не имеет и не может иметь нулевого значения, то вышеуказанная трактовка параметра  $a$  не имеет смысла.

Оценка тесноты связи между признаками осуществляется с помощью коэффициента линейной парной корреляции -  $r_{x,y}$ . Он может быть рассчитан по

$$r_{x,y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

формуле: . Кроме того, коэффициент линейной парной корреляции может быть определен через

$$r_{x,y} = b \frac{\sigma_x}{\sigma_y}$$

коэффициент регрессии  $b$ :

Область допустимых значений линейного коэффициента парной корреляции от  $-1$  до  $+1$ . Знак коэффициента корреляции указывает направление связи. Если  $r_{x,y} > 0$ , то связь прямая; если  $r_{x,y} < 0$ , то связь обратная.

Если данный коэффициент по модулю близок к единице, то связь между признаками может быть интерпретирована как довольно тесная линейная. Если его модуль равен единице  $|r_{x,y}| = 1$ , то связь между признаками функциональная линейная. Если признаки  $x$  и  $y$  линейно независимы, то  $r_{x,y}$  близок к  $0$ .

Для расчета  $r_{x,y}$  можно использовать также таблицу 1.

Таблица 1



	i	x <sub>i</sub>	y <sub>i</sub>	x <sub>i</sub> ·y <sub>i</sub>	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	1	x <sub>1</sub>	y <sub>1</sub>	x <sub>1</sub> ·y <sub>1</sub>	$(x_1 - \bar{x})^2$	$(y_1 - \bar{y})^2$
2	2	x <sub>2</sub>	y <sub>2</sub>	x <sub>2</sub> ·y <sub>2</sub>	$(x_2 - \bar{x})^2$	$(y_2 - \bar{y})^2$
...						
n	n	x <sub>n</sub>	y <sub>n</sub>	x <sub>n</sub> ·y <sub>n</sub>	$(x_n - \bar{x})^2$	$(y_n - \bar{y})^2$
Сумма по столбцу	Σ	Σ	Σ	Σ	$\sum (x_i - \bar{x})^2$	$\sum (y_i - \bar{y})^2$
Среднее значение	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{y} = \frac{\sum y_i}{n}$	$\bar{x} \cdot \bar{y} = \frac{\sum x_i \cdot y_i}{n}$	$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$	$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$	

Для оценки качества полученного уравнения регрессии рассчитывают теоретический коэффициент детерминации – R<sup>2</sup><sub>yx</sub>:

$$R_{yx}^2 = \frac{\delta^2}{\sigma_y^2} = \frac{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\varepsilon^2}{\sigma_y^2} = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

,

где δ<sup>2</sup> – объясненная уравнением регрессии дисперсия y;

$\varepsilon^2$ - остаточная (необъясненная уравнением регрессии) дисперсия  $y$ ;

$\sigma^2y$  - общая (полная) дисперсия  $y$ .

Коэффициент детерминации характеризует долю вариации (дисперсии) результативного признака  $y$ , объясняемую регрессией ( $a$ , следовательно, и фактором  $x$ ), в общей вариации (дисперсии)  $y$ . Коэффициент детерминации  $R^2_{yx}$  принимает значения от 0 до 1. Соответственно величина  $1-R^2_{yx}$  характеризует долю дисперсии  $y$ , вызванную влиянием прочих неучтенных в модели факторов и ошибками спецификации.

При парной линейной регрессии  $R^2_{yx} = r^2_{yx}$ .

Оценка статистической значимости параметров уравнения регрессии.

С помощью МНК мы получили лишь оценки параметров уравнения регрессии, которые характерны для конкретного статистического наблюдения (конкретного набора значений  $x$  и  $y$ ). Если оценку параметров произвести по данным другого статистического наблюдения (другому набору значений  $x$  и  $y$ ), то получим

другие численные значения  $\tilde{a}$ ,  $\tilde{b}$ . Мы предполагаем, что все эти наборы значений  $x$  и  $y$  извлечены из одной и той же генеральной совокупности. Чтобы проверить, значимы ли параметры, т.е. значимо ли они отличаются от нуля для генеральной совокупности используют статистические методы проверки гипотез.

В качестве основной (нулевой) гипотезы выдвигают гипотезу о незначимом отличии от нуля параметра или статистической характеристики в генеральной

совокупности. Наряду с основной (проверяемой) гипотезой выдвигают альтернативную (конкурирующую) гипотезу о неравенстве нулю параметра или статистической характеристики в генеральной совокупности. В случае если основная гипотеза окажется неверной, мы принимаем альтернативную. Для проверки этой гипотезы используется t-критерий Стьюдента.

Найденное по данным наблюдений значение t-критерия (его еще называют наблюдаемым или фактическим) сравнивается с табличным (критическим) значением, определяемым по таблицам распределения Стьюдента (которые обычно приводятся в конце учебников и практикумов по статистике или эконометрике). Табличное значение определяется в зависимости от уровня значимости ( $\alpha$ ) и числа степеней свободы, которое в случае линейной парной регрессии равно  $(n-2)$ ,  $n$ -число наблюдений.

Если фактическое значение t-критерия больше табличного (по модулю), то основную гипотезу отвергают и считают, что с вероятностью  $(1-\alpha)$  параметр или статистическая характеристика в генеральной совокупности значимо отличается от нуля.

Если фактическое значение t-критерия меньше табличного (по модулю), то нет оснований отвергать основную гипотезу, т.е. параметр или статистическая характеристика в генеральной совокупности незначимо отличается от нуля при уровне значимости  $\alpha$ .

Для параметра  $b$  критерий проверки имеет вид:

$$t_{(b=0)} = \frac{\tilde{b}}{\mu_{\tilde{b}}},$$

где  $\tilde{b}$  - оценка коэффициента регрессии, полученная по наблюдаемым данным;

$\mu_{\tilde{b}}$  - стандартная ошибка коэффициента регрессии.

Для линейного парного уравнения регрессии стандартная ошибка коэффициента вычисляется по формуле:

$$\mu_{\tilde{b}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Числитель в этой формуле может быть рассчитан через коэффициент детерминации и общую дисперсию

признака-результата:  $\sum_{i=1}^n (y_i - \tilde{y}_i)^2 = n \cdot (1 - R_{yx}^2) \cdot \sigma_y^2$ .

Для параметра  $a$  критерий проверки гипотезы о незначимом отличии его от нуля имеет вид:

$$t_{(a=0)} = \frac{\tilde{a}}{\mu_{\tilde{a}}},$$

где  $\tilde{a}$  - оценка параметра регрессии, полученная по наблюдаемым данным;

$\mu_{\tilde{a}}$  – стандартная ошибка параметра а.

Для линейного парного уравнения регрессии:

$$\mu_{\tilde{a}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2 \sum_{i=1}^n x_i^2}{n(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Для проверки гипотезы о незначимом отличии от нуля коэффициента линейной парной корреляции в генеральной совокупности используют следующий критерий:

$$t_{(r=0)} = \frac{r_{yx}}{\mu_r}$$

, где  $r_{yx}$  – оценка коэффициента корреляции, полученная по наблюдаемым данным;  $\mu_r$  – стандартная ошибка коэффициента корреляции  $r_{yx}$ .

Для линейного парного уравнения регрессии:

$$\mu_r = \sqrt{\frac{(1 - r_{yx}^2)}{(n - 2)}}$$

В парной линейной регрессии между наблюдаемыми значениями критериев существует взаимосвязь:  $t(b=0) = t(r=0)$ .

Прогноз ожидаемого значения результативного признака у по линейному парному уравнению регрессии.

Пусть требуется оценить значение признака-результата для заданного значения признака-фактора (х<sub>р</sub>). Прогнозируемое значение признака-результата с

доверительной вероятностью равной  $(1-\alpha)$  принадлежит интервалу прогноза:

$$(\tilde{y}^p - t \cdot \mu_p; \tilde{y}^p + t \cdot \mu_p),$$

где  $\tilde{y}^p$  - точечный прогноз;

$t$  – коэффициент доверия, определяемый по таблицам распределения Стьюдента в зависимости от уровня значимости  $\alpha$  и числа степеней свободы  $(n-2)$ ;

$\mu_p$  - средняя ошибка прогноза.

Точечный прогноз рассчитывается по линейному уравнению регрессии, как:  $\tilde{y}^p = \tilde{a} + \tilde{b} \cdot x^p$ .

Средняя ошибка прогноза определяется по формуле:

$$\mu_p = \sqrt{\frac{\sum (y_i - \tilde{y}_i)^2}{n-2} \left( 1 + \frac{1}{n} + \frac{(x^p - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)}.$$

### Задание № 1

На основе данных, приведенных в Приложении 1 и соответствующих Вашему варианту (таблица 2), требуется:

Рассчитать коэффициент линейной парной корреляции и построить уравнение линейной парной регрессии одного признака от другого. Один из признаков, соответствующих Вашему варианту, будет играть роль факторного ( $x$ ), другой – результативного ( $y$ ). Причинно-следственные связи между признаками установить самим на основе экономического анализа. Пояснить смысл параметров уравнения.

Определить теоретический коэффициент детерминации и остаточную (необъясненную уравнением регрессии) дисперсию. Сделать вывод.

Оценить статистическую значимость уравнения регрессии в целом на пятипроцентном уровне с помощью F-критерия Фишера. Сделать вывод.

Выполнить прогноз ожидаемого значения признака-результата у при прогнозном значении признака-фактора  $x$ , составляющим 105% от среднего уровня  $x$ . Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал с вероятностью 0,95.

Таблица 2

а р и а н т	омер начал ьного набл юден ия	омер конеч ного набл юден ия	омер приз нако в из прил . 1	ар иа нт	омер начал ьного набл юден ия	омер конеч ного набл юден ия	Номер
							признаков и прил. 1
							8
1		0	,2	1	6	5	1,3
2		0	,4	2	6	5	4,5
3		1	,3	3	7	6	1,4
4		1	,5	4	7	6	2,5

5		2	,4	5	8	7	1,5
6		2	,5	6	8	7	2,3
7		3	,5	7	9	8	1,2
8		3	,3	8	9	8	3,4
9		4	,2	9	0	9	1,3
0		4	,4	0	0	9	4,5
1		5	,3	1	1	0	1,4
2		5	,5	2	1	0	2,5
3		6	,4	3	2	1	1,5
4		6	,5	4	2	1	2,3
5		7	,5	5	3	2	1,2
6		7	,3	6	3	2	3,4
7		8	,2	7	4	3	1,3
8		8	,4	8	4	3	4,5



9	0	9	,3	9	5	4	1,4
0	0	9	,5	0	5	4	2,5
1	1	0	,4	1	6	5	1,5
2	1	0	,5	2	6	5	2,3
3	2	1	,5	3	7	6	1,2
4	2	1	,3	4	7	6	3,4
5	3	2	,2	5	8	7	1,3
6	3	2	,4	6	8	7	4,5
7	4	3	,3	7	9	8	1,4
8	4	3	,5	8	9	8	2,5
9	5	4	,4	9	0	9	1,5
0	5	4	,5	0	0	9	2,3
1	6	5	,5	1	1	0	1,2
2	6	5	,3	2	1	0	3,4

3	7	6	,2	3	2	1	1,3
4	7	6	,4	4	2	1	4,5
5	8	7	,3	5	3	2	1,4
6	8	7	,5	6	3	2	2,5
7	9	8	,4	7	4	3	1,5
8	9	8	,5	8	4	3	2,3
9	0	9	,5	9	5	4	1,2
0	0	9	,3	0	5	4	3,4
1	1	0	,2	1	6	5	1,3
2	1	0	,4	2	6	5	4,5
3	2	1	,3	3	7	6	1,4
4	2	1	,5	4	7	6	2,5
5	3	2	,4	5	8	7	1,5
6	3	2	,5	6	8	7	2,3

7	4	3	,5	7	9	8	1,2
8	4	3	,3	8	9	8	3,4
9	5	4	,2	9	0	9	1,3
0	5	4	,4		0	9	4,5

## Тема 2. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели, который в свою очередь включает 2 круга вопросов: отбор факторов и выбор уравнения регрессии.

Отбор факторов обычно осуществляется в два этапа:

1) теоретический анализ взаимосвязи результата и круга факторов, которые оказывают на него существенное влияние;

2) количественная оценка взаимосвязи факторов с результатом. При линейной форме связи между признаками данный этап сводится к анализу корреляционной матрицы (матрицы парных линейных коэффициентов корреляции):

$$\begin{array}{ccccccc} r_{y,y} & r_{y,x_1} & r_{y,x_2} & \dots & r_{y,x_m} \\ r_{x_1,y} & r_{x_1,x_1} & r_{x_1,x_2} & \dots & r_{x_1,x_m} \\ \dots & & & & \end{array}$$

$$r_{x_m,y} \quad r_{x_m,x_1} \quad r_{x_m,x_2} \quad \dots \quad r_{x_m,x_m}$$

где  $r_{y,x_j}$  – линейный парный коэффициент корреляции, измеряющий тесноту связи между признаками  $y$  и  $x_j$   $j=1;m$ ,  $m$  -число факторов.

$r_{x_j,x_k}$  – линейный парный коэффициент корреляции, измеряющий тесноту связи между признаками  $x_j$  и  $x_k$   $j,k=1;m$ .

Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1. Они должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не

имеющий количественного измерения, то ему нужно придать количественную определенность (например, в модели урожайности качество почвы задается в виде баллов).

2. Каждый фактор должен быть достаточно тесно связан с результатом (т.е. коэффициент парной линейной корреляции между фактором и результатом должен быть существенным).

3. Факторы не должны быть сильно коррелированы друг с другом, тем более находиться в строгой функциональной связи (т.е. они не должны быть интеркоррелированы).

Разновидностью интеркоррелированности факторов является мультиколлинеарность - тесная линейная связь между факторами.

Мультиколлинеарность может привести к нежелательным последствиям:

1) оценки параметров становятся ненадежными. Они обнаруживают большие стандартные ошибки. С изменением объема наблюдений оценки меняются (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

2) затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл;

3) становится невозможным определить изолированное влияние факторов на результирующий показатель.

Мультиколлинеарность имеет место, если определитель матрицы межфакторной корреляции близок к нулю:

$$\text{Det}|R| = \begin{vmatrix} r_{x1x1} & r_{x2x1} & r_{x3x1} \\ r_{x1x2} & r_{x2x2} & r_{x3x2} \\ r_{x1x3} & r_{x2x3} & r_{x3x3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

Если же определитель матрицы межфакторной корреляции близок к единице, то мультиколлинеарности нет. Существуют различные подходы преодоления сильной межфакторной корреляции. Простейший из них – исключение из модели фактора (или факторов), в наибольшей степени ответственных за мультиколлинеарность при условии, что качество модели при этом пострадает незначительно (а именно, теоретический коэффициент детерминации  $R^2y(x1...xm)$  снизится незначительно).

Определение факторов, ответственных за мультиколлинеарность, может быть основано на анализе матрицы межфакторной корреляции. При этом определяют пару признаков-факторов, которые сильнее всего связаны между собой (коэффициент линейной парной корреляции максимален по модулю). Из этой пары в наибольшей степени ответственным за мультиколлинеарность будет тот признак, который теснее связан с другими факторами модели (имеет более высокие по модулю значения коэффициентов парной линейной корреляции).

Еще один способ определения факторов, ответственных за мультиколлинеарность основан на вычислении коэффициентов множественной детерминации

( $R^2_{x_j}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m)$ ), показывающего зависимость фактора  $x_j$  от других факторов модели  $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m$ . Чем ближе значение коэффициента множественной детерминации к единице, тем больше ответственность за мультиколлинеарность фактора, выступающего в роли зависимой переменной. Сравнивая между собой коэффициенты множественной детерминации для различных факторов можно проранжировать переменные по степени ответственности за мультиколлинеарность.

При выборе формы уравнения множественной регрессии предпочтение отдается линейной функции:

$$y_i = a + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + \dots + b_m \cdot x_{mi} + u_i$$

в виду четкой интерпретации параметров.

Данное уравнение регрессии называют уравнением регрессии в естественном (натуральном) масштабе. Коэффициент регрессии  $b_j$  при факторе  $x_j$  называют условно-чистым коэффициентом регрессии. Он измеряет среднее по совокупности отклонение признака-результата от его средней величины при отклонении признака-фактора  $x_j$  на единицу, при условии, что все прочие факторы модели не изменяются (зафиксированы на своих средних уровнях).

Если не делать предположения о значениях прочих факторов, входящих в модель, то это означало бы, что каждый из них при изменении  $x_j$  также изменялся бы (так как факторы связаны между собой), и своими изменениями оказывали бы влияние на признак-результат.

Расчет параметров уравнения линейной множественной регрессии.

Параметры уравнения множественной регрессии можно оценить методом наименьших квадратов, составив и решив систему нормальных линейных уравнений.

Кроме того, для линейной множественной регрессии существует другой способ реализации МНК при оценке параметров - через  $\beta$ -коэффициенты (через параметры уравнения регрессии в стандартных масштабах).

Модель регрессии в стандартном масштабе предполагает, что все значения исследуемых признаков переводятся в стандарты (стандартизованные значения) по формулам:

$$t_{x_j} = \frac{x_{j_i} - \bar{x}_j}{\sigma_{x_j}}, \quad j=1;m,$$

где  $x_{j_i}$  - значение переменной  $x_j$  в  $i$ -ом наблюдении.

$$t_{y_i} = \frac{y_i - \bar{y}}{\sigma_y}.$$

Таким образом, начало отсчета каждой стандартизованной переменной совмещается с ее средним значением, а в качестве единицы изменения принимается ее среднее квадратическое отклонение  $\sigma$ . Если связь между переменными в естественном масштабе линейная, то изменение начала отсчета и единицы измерения этого свойства не нарушат, так что и стандартизованные переменные будут связаны линейным соотношением:



$$\tilde{t}_y = \sum_{j=1}^m \beta_j \cdot t_{xj}$$

Для оценки  $\beta$ -коэффициентов применим МНК. При этом система нормальных уравнений будет иметь вид:

$$r_{x1y} = \beta_1 + r_{x1x2} \cdot \beta_2 + \dots + r_{x1xm} \cdot \beta_m$$

$$r_{x2y} = r_{x2x1} \cdot \beta_1 + \beta_2 + \dots + r_{x2xm} \cdot \beta_m$$

...

$$r_{xmy} = r_{xmx1} \cdot \beta_1 + r_{xmx2} \cdot \beta_2 + \dots + \beta_m$$

Найденные из данной системы  $\beta$ -коэффициенты позволяют определить значения коэффициентов в регрессии в естественном масштабе по формулам:

$$\tilde{b}_j = \beta_j \cdot \frac{\sigma_y}{\sigma_x}, \quad j=1; m; \quad \tilde{a} = \bar{y} - \sum_{j=1}^m \tilde{b}_j \cdot \bar{x}_j$$

Показатели тесноты связи факторов с результатом.

Если факторные признаки различны по своей сущности и (или) имеют различные единицы измерения, то коэффициенты регрессии  $b_j$  при разных факторах являются несопоставимыми. Поэтому уравнение регрессии дополняют соизмеримыми показателями тесноты связи фактора с результатом, позволяющими ранжировать факторы по силе влияния на результат. К таким показателям тесноты связи относят: частные коэффициенты эластичности,  $\beta$ -коэффициенты, частные коэффициенты корреляции.

Частные коэффициенты эластичности  $\mathcal{E}_j$

$$\mathcal{E}_j = \frac{\partial y}{\partial x_j} \cdot \frac{\overline{x_j}}{\overline{y_{x_1, \dots, x_m}}}$$

рассчитываются по формуле:

Частный коэффициент эластичности показывают, на сколько процентов в среднем изменяется признак-результат  $y$  с изменением признака-фактора  $x_j$  на один процент от своего среднего уровня при фиксированном положении других факторов модели. В случае линейной зависимости  $\mathcal{E}_j$  рассчитываются по формуле:

$$\mathcal{E}_j = \tilde{b}_j \cdot \frac{\overline{x_j}}{\overline{y_{x_1, \dots, x_m}}}$$
, где  $\tilde{b}_j$  – оценка коэффициента регрессии при  $j$ -ом факторе.

Стандартизированные частные коэффициенты регрессии -  $\beta$ -коэффициенты ( $\beta_j$ ) показывают, на какую часть своего среднего квадратического отклонения  $\sigma_y$  изменится признак-результат  $y$  с изменением соответствующего фактора  $x_j$  на величину своего среднего квадратического отклонения ( $\sigma_{x_j}$ ) при неизменном влиянии прочих факторов (входящих в уравнение).

По коэффициентам эластичности и  $\beta$ -коэффициентам могут быть сделаны противоположные выводы. Причины этого: а) вариация одного фактора очень велика; б) разнонаправленное воздействие факторов на результат.

Коэффициент  $\beta_j$  может также интерпретироваться как показатель прямого (непосредственного) влияния  $j$ -ого фактора ( $x_j$ ) на результат ( $y$ ). Во множественной регрессии

$j$ -ый фактор оказывает не только прямое, но и косвенное (опосредованное) влияние на результат (т.е. влияние через другие факторы модели). Косвенное влияние измеряется

величиной:  $\sum_{i=1, \dots, j-1, j+1, \dots, m} \beta_i \cdot r_{x_j, x_i}$ , где  $m$ - число факторов в модели. Полное влияние  $j$ -ого фактора на результат равно сумме прямого и косвенного влияний измеряет коэффициент линейной парной корреляции данного фактора и результата –  $r_{x_j, y}$ .

Коэффициент частной корреляции измеряет «чистое» влияние фактора на результат при устранении воздействия прочих факторов модели.

Для расчета частных коэффициентов корреляции могут быть использованы парные коэффициенты корреляции.

Для случая зависимости  $y$  от двух факторов можно вычислить 2 коэффициента частной корреляции:

$$r_{yx1/x2} = \frac{r_{x1y} - r_{x2y} \cdot r_{x1x2}}{\sqrt{(1 - r_{x1x2}^2)(1 - r_{x2y}^2)}},$$

(фактор  $x_2$  фиксирован).

$$r_{yx2/x1} = \frac{r_{x2y} - r_{x1y} \cdot r_{x1x2}}{\sqrt{(1 - r_{x1x2}^2)(1 - r_{x1y}^2)}}$$

(фактор  $x_1$  фиксирован).

Это коэффициенты частной корреляции 1-ого порядка (порядок определяется числом факторов, влияние которых устраняется).

Частные коэффициенты корреляции, рассчитанные по таким формулам изменяются от  $-1$  до  $+1$ . Они

используются не только для ранжирования факторов модели по степени влияния на результат, но и также для отсева факторов. При малых значениях  $\beta_{x1}, \beta_{x2}, \dots, \beta_{xm-1}$  нет смысла вводить в уравнение  $m$ -ый фактор, т.к. его чистое влияние на результат несущественно.

Коэффициенты множественной детерминации и корреляции характеризуют совместное влияние всех факторов на результат.

По аналогии с парной регрессией можно определить долю вариации результата, объясненной вариацией включенных в модель факторов ( $\delta^2$ ), в его общей вариации ( $\sigma^2_y$ ). Ее количественная характеристика – теоретический множественный коэффициент детерминации ( $R^2_{y(x1, \dots, xm)}$ ). Для линейного уравнения регрессии данный показатель может быть рассчитан через  $\beta$ -коэффициенты, как:

$$R^2_{y(x1, \dots, xm)} = \sum_{j=1}^m \beta_j \cdot r_{y xj}$$

$$R_{y(x1, \dots, xm)} = \sqrt{R^2_{y(x1, \dots, xm)}}$$

- коэффициент множественной корреляции. Он принимает значения от 0 до 1 (в отличие от парного коэффициента корреляции, который может принимать отрицательные значения). Поэтому  $R$  не может быть использован для интерпретации направления связи. Чем плотнее фактические значения  $y_i$  располагаются относительно линии регрессии, тем меньше остаточная дисперсия и, следовательно, больше величина  $R_{y(x1, \dots, xm)}$ . Таким образом, при значении  $R$  близком к 1, уравнение регрессии лучше описывает фактические

данные и факторы сильнее влияют на результат. При значении R близком к 0 уравнение регрессии плохо описывает фактические данные и факторы оказывают слабое воздействие на результат.

Оценка значимости полученного уравнения множественной регрессии.

Оценка значимости уравнения множественной регрессии осуществляется путем проверки гипотезы о равенстве нулю коэффициент детерминации рассчитанного по данным генеральной совокупности:  $R_{y(x_1, \dots, x_m)}^{2(\varepsilon)} = 0$  или  $b_1 = b_2 = \dots = b_m = 0$  (гипотеза о незначимости уравнения регрессии, рассчитанного по данным генеральной совокупности).

Для ее проверки используют F-критерий Фишера.

При этом вычисляют фактическое (наблюдаемое) значение F-критерия, через коэффициент детерминации  $R_{y(x_1, \dots, x_m)}^2$ , рассчитанный по данным конкретного наблюдения:

$$F = \frac{R_{y(x_1, \dots, x_m)}^2}{1 - R_{y(x_1, \dots, x_m)}^2} \cdot \frac{n - h}{h - 1}, \text{ где } n - \text{число наблюдений;}$$

$h$  – число оцениваемых параметров (в случае двухфакторной линейной регрессии  $h=3$ ).

По таблицам распределения Фишера-Снедеккера находят критическое значение F-критерия ( $F_{кр}$ ). Для этого задаются уровнем значимости  $\alpha$  (обычно его берут равным 0,05) и двумя числами степеней свободы  $k_1=h-1$  и  $k_2=n-h$ .

Сравнивают фактическое значение F-критерия ( $F_{набл}$ ) с табличным  $F_{кр}(\alpha; k_1; k_2)$ . Если

$F_{\text{набл}} < F_{\text{кр}}(\alpha; k_1; k_2)$ , то гипотезу о незначимости уравнения регрессии не отвергают. Если  $F_{\text{набл}} > F_{\text{кр}}(\alpha; k_1; k_2)$ , то выдвинутую гипотезу отвергают и принимают альтернативную гипотезу о статистической значимости уравнения регрессии.

## Задание № 2

На основе данных, приведенных в Приложении и соответствующих Вашему варианту (таблица 2), требуется:

Построить уравнение множественной регрессии. При этом признак-результат и один из факторов остаются теми же, что и в первом задании. Выберите дополнительно еще один фактор из приложения 1 (границы наблюдения должны совпадать с границами наблюдения признака-результата, соответствующего Вашему варианту). При выборе фактора нужно руководствоваться его экономическим содержанием или другими подходами. Пояснить смысл параметров уравнения.

Рассчитать частные коэффициенты эластичности. Сделать вывод.

Определить стандартизованные коэффициенты регрессии ( $\beta$ -коэффициенты). Сделать вывод.

Определить парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделать выводы.

Оценить значимость параметров уравнения регрессии с помощью  $t$ -критерия Стьюдента, а также значимость уравнения регрессии в целом с помощью общего  $F$ -критерия Фишера. Предложить окончательную модель (уравнение регрессии). Сделать выводы.