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): Учёт, анализ и аудит
Форма обучения: заочная, ускоренная
Квалификация (степень) выпускника: Бакалавр

Новороссийск 2018

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. Попытайтесь 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 результативного 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).

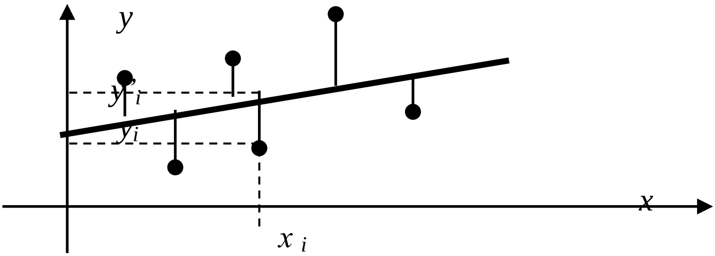
Задача оценивания параметров линейного парного уравнения методом наименьших квадратов состоит в следующем:

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

Формально критерий МНК можно записать так:

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

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:

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

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

В результате получим систему из 2-ух нормальных линейных уравнений:
$$\begin{cases} \sum_{i=1}^{n} y_i = \tilde{a} \cdot n + \tilde{b} \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i \cdot x_i = \tilde{a} \sum_{i=1}^{n} x_i + \tilde{b} \sum_{i=1}^{n} x_i^2 \end{cases}$$

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

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

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

Формально значение параметра *a* – среднее значение у при х равном нулю. Если признак-фактор не имеет и не может иметь нулевого значения, то вышеуказанная трактовка параметра а не имеет смысла.

Оценка тесноты связи между признаками осуществляется с помощью коэффициента линейной парной корреляции - $r_{x,y}$. Он может быть рассчитан по формуле: $r_{x,y} = \dfrac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}$. Кроме того, коэффициент линейной парной корреляции может быть определен через коэффициент регрессии b: $r_{x,y} = b\dfrac{\sigma_x}{\sigma_y}$.

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

	x_i	$_i$	$x_i \cdot y_i$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$
1	x_1	y_1	$X_1 \cdot Y_1$	$(x_1 - \overline{x})^2$	$(y_1 - \overline{y})^2$
2	x_2	y_2	$X_2 \cdot Y_2$	$(x_2 - \overline{x})^2$	$(y_2 - \overline{y})^2$
...					
n	x	y	$X_N \cdot$	$(x_n - \overline{x})^2$	$(y_m - \overline{y})^2$

	n	n	$\boldsymbol{Y_N}$		
Сумма по столбцу	$\sum x$	$\sum y$	$\sum x\cdot y$	$\sum(x_i-\bar{x})^2$	$\sum(y_i-\bar{y})^2$
Среднее значение	$\bar{x}=\dfrac{\sum x_i}{n}$	$\bar{y}=\dfrac{\sum y_i}{n}$	$\overline{x\cdot y}=\dfrac{\sum x_i\cdot y_i}{n}$	$\sigma_x^2=\dfrac{\sum(x_i-\bar{x})^2}{n}$	$\sigma_y^2=\dfrac{\sum(y_i-\bar{y})^2}{n}$

Для оценки качества полученного уравнения регрессии рассчитывают теоретический коэффициент детерминации – R^2_{yx}:

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

где δ^2 – объясненная уравнением регрессии дисперсия y;

ε^2- остаточная (необъясненная уравнением регрессии) дисперсия y;

σ^2_y - общая (полная) дисперсия y.

Коэффициент детерминации характеризует долю вариации (дисперсии) результативного признака y, объясняемую регрессией (а, 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 (α) 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-α) 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 α.

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 *a*.

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; μ_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 (x^p). Прогнозируемое значение признака-результата с доверительной вероятностью равной (1-α) 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 α и числа степеней свободы (n-2);

μ_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:

1. 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.
2. Определить теоретический коэффициент детерминации и остаточную (необъясненную уравнением регрессии) дисперсию. Сделать вывод.
3. Оценить статистическую значимость уравнения регрессии в целом на пятипроцентном уровне с помощью F-критерия Фишера. Сделать вывод.
4. Выполнить прогноз ожидаемого значения признака-результата y 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

ариант	омер начального наблюдения	омер конечного наблюдения	омер признаков из прил. 1	ариант	омер начального наблюдения	омер конечного наблюдения	Номер признаков из 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):

$$\begin{matrix} r_{y,y} & r_{y,x1} & r_{yx2} & \ldots & r_{y,xm} \\ r_{x1,y} & r_{x1,x2} & r_{x2x2} & \ldots & r_{x2,xm} \\ \ldots\ldots & & & & \\ r_{xm,y} & r_{xm,x1} & r_{xm,x2} & \ldots & r_{xm,xm} \end{matrix}$$

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

$r_{xj,xk}$ – линейный парный коэффициент корреляции, измеряющий тесноту связи между признаками xj и xk 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 коррелированны; параметры линейной регрессии теряют экономический смысл;

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

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

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

Если же определитель матрицы межфакторной корреляции близок к единице, то мульткоплинеарности 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 несущественно (а именно, теоретический коэффициент детерминации -$R^2_{y(x1...xm)}$ снизится несущественно).

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 ($R^2_{xj(x1,...,xj-1,xj+1,...,xm)}$), показывающего зависимость фактора *xj* от других факторов модели *x1,...,xj-1, xj+1,...,xm*. 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:

$$y_i = a + b_1 \cdot x1_i + b_2 \cdot x2_i + \ldots + b_m \cdot xm_i + u_i$$

в виду четкой интерпретации параметров.

Данное уравнение регрессии называют уравнением регрессии в естественном (натуральном) масштабе. Коэффициент регрессии b_j при факторе *xj* называют *условно-чистым коэффициентом регрессии*. Он измеряет среднее по совокупности отклонение признака-результата от его средней величины при отклонении признака-фактора *xj* 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 *xj* 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 β-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:

$$t_{xj_i} = \frac{xj_i - \overline{xj}}{\sigma_{xj}}, \quad j=1;m,$$

где xj_i - значение переменной xj_i в i-ом наблюдении.

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

Таким образом, начало отсчета каждой стандартизованной переменной совмещается с ее средним значением, а в качестве единицы изменения принимается ее среднее квадратическое отклонение σ. 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 β-коэффциентов применим МНК. При этом система нормальных уравнений будет иметь вид:

$r_{x1y}=\beta 1+r_{x1x2}\cdot\beta 2+\ldots+r_{x1xm}\cdot\beta m$

$r_{x2y}=r_{x2x1}\cdot\beta 1+\beta 2+\ldots+r_{x2xm}\cdot\beta m$

…

$r_{xmy}=r_{xmx1}\cdot\beta 1+r_{xmx2}\cdot\beta 2+\ldots+\beta m$

Найденные из данной системы β–коэффициенты позволяют определить значения коэффициентов в регрессии в естественном масштабе по формулам:

$$\tilde{b}_j = \beta j \cdot \frac{\sigma_y}{\sigma_x},\ j=1;m; \quad \tilde{a} = \overline{y} - \sum_{j=1}^{m} \tilde{b}_j \cdot \overline{xj}\,.$$

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, β–коэффициенты, частные коэффициенты корреляции.

Частные коэффициенты эластичности $Э_j$ рассчитываются по формуле: $Э_j = \frac{\partial y}{\partial x_j} \cdot \frac{\overline{x_j}}{\overline{\overline{y_{x1,\ldots,xm}}}}$. Частный коэффициент эластичности показывают, на сколько процентов в среднем изменяется признак-результат у с изменением признака-фактора *xj* на один процент от своего среднего уровня при фиксированном положении других факторов модели. В случае линейной зависимости $Э_j$ рассчитываются по формуле: $Э_j = \tilde{b}_j \cdot \frac{\overline{x}_j}{\overline{\overline{y}}_{x1,\ldots,xm}}$, где $\tilde{b}_j$ – оценка коэффициента регрессии при j–ом факторе.

Стандартизированные частные коэффициенты регрессии - β-коэффициенты (β_j) показывают, на какую часть своего среднего квадратического отклонения σ_y изменится признак-результат y с изменением соответствующего фактора xj на величину своего среднего квадратического отклонения (σ_{xj}) при неизменном влиянии прочих факторов (входящих в уравнение).

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

Коэффициент β_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: $\sum\limits_{i=1,\ldots,j-1,j+1,\ldots,m} \beta_i \cdot r_{xj,xi}$, где m- число факторов в модели. Полное влияние j-ого фактора на результат равное сумме прямого и косвенного влияний измеряет коэффициент линейной парной корреляции данного фактора и результата – $r_{xj,y}$.

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:

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

(фактор x2 фиксирован).

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

(фактор x1 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 $r_{yxm/x1,x2...xm-1}$ 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 (δ^2), в его общей вариации (σ^2_y). Ее количественная характеристика – теоретический множественный коэффициент детерминации ($R^2_{y(x1,...,xm)}$). Для линейного уравнения регрессии данный показатель может быть рассчитан через β-коэффициенты, как:

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

$R_{y(x1,...,xm)} = \sqrt{R^2_{y(x1,...,xm)}}$ - коэффициент множественной корреляции. Он принимает значения от 0 до 1 (в отличии 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 $R_{y(x1,...,xm)}$. 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: $R^{2(г)}_{y(x1,..,xm)} = 0$ или $b_1 = b_2 = \ldots = b_m = 0$ (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 $R^2_{y(x1,...,xm)}$, рассчитанный по данным конкретного наблюдения:

$F = \frac{R^2_{y(x1,...,xm)}}{1 - R^2_{y(x1,...,xm)}} \cdot \frac{n-h}{h-1}$, где *n*-число наблюдений; *h* – число оцениваемых параметров (в случае двухфакторной линейной регрессии *h*=3).

По таблицам распределения Фишера-Снедоккора находят критическое значение *F*-критерия ($F_{кр}$). Для этого задаются уровнем значимости α (обычно его берут равным 0,05) и двумя числами степеней свободы $k1=h-1$ и $k2=n-h$.

Сравнивают фактическое значение F-критерия ($F_{набл}$) с табличным $F_{кр}(\alpha;k1;k2)$. Если $F_{набл}<F_{кр}(\alpha;k1;k2)$, то гипотезу о незначимости уравнения регрессии не отвергают. Если $F_{набл}>F_{кр}(\alpha;k1;k2)$, 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:

1. 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.
2. Рассчитать частные коэффициенты эластичности. Сделать вывод.
3. Определить стандартизованные коэффициенты регрессии (β-коэффициенты). Сделать вывод.
4. Определить парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделать выводы.
5. 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.