Лабораторная работа № 1 (2 часа)
Метод наименьших квадратов (МНК). Случай двух переменных

З а д а н и е. С помощью МНК по выборке (x_t, y_t) $(t = 1,\ldots,n)$ оценить зависимость переменной $y = (y_1,\ldots, y_n)$ от переменной $x = (x_1,\ldots, x_n)$ считая, что она линейна, т.е. имеет вид:

$$\hat{y} = \alpha + \beta x.$$

Напомним расчетные формулы МНК

$$\hat{\beta} = \frac{n\sum_{t=1}^{n} x_t y_t - \sum_{t=1}^{n} x_t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} x_t^2 - \left(\sum_{t=1}^{n} x_t\right)^2},\ \hat{\alpha} = \frac{1}{n}\sum_{t=1}^{n} y_t - \hat{\beta}\frac{1}{n}\sum_{t=1}^{n} x_t\ . \qquad (*)$$

При вычислениях использовать Excel.

П р и м е р. Имеется набор данных

x_t	2	4	6	7	9
y_t	5	10	10	15	20

Предположим, что переменная ***y*** линейно зависит от переменной ***x***: ***y*** = $=\alpha+\beta\boldsymbol{x}$. Требуется с помощью МНК найти приближенные значения параметров α и β этой зависимости.

Р е ш е н и е. Для удобства оформим вычисления в виде таблицы:

№	x_t	y_t	$x_t y_t$	x_t^2
1	2	5	10	4
2	4	10	40	16
3	6	10	60	36
4	7	15	105	49
5	9	20	180	81
∑	**28**	**60**	**395**	**184**

Применяя формулы (*), получаем:

$$\hat{\beta} = \frac{5\cdot 395 - 28\cdot 60}{5\cdot 184 - 28^2} = \frac{295}{136} \approx 2{,}17;\ \hat{\alpha} = \frac{60}{5} - 2{,}17\cdot\frac{28}{5} \approx -0{,}15.$$

Итак, (с точностью до двух знаков после запятой) искомая зависимость описывается прямой $y = - 0{,}15 + 2{,}17x$.

Лабораторная работа № 2 (6 часов)
Классическая модель парной регрессии. Основные статистики

З а д а н и е. Оценить модель парной регрессии по следующим пунктам.

1. С помощью МНК по выборке (x_t, y_t) $(t = 1,2,\ldots,n)$ оценить модель регрессии переменной y на переменную x (x и y – конкретные экономические показатели) и дать экономическую интерпретацию коэффициентам модели.

2. Найти стандартную ошибку и стандартные ошибки коэффициентов регрессии.

3. Проверить на 5%-ю значимость коэффициенты регрессии.

4. Построить 955-е доверительные интервалы для коэффициентов регрессии и по их виду сделать вывод о значимости коэффициентов.

5. Найти коэффициент детерминации модели и сделать вывод о качестве аппроксимации данных.

6. Проверить на 5%-ю значимость регрессию в целом.

П р и м е р. Имеются выборочные данные по пяти группам предприятий (т.е. выборка объема $n = 5$) за отчетный период (y – себестоимость единицы продукции, x – выпуск):

x	1	3	5	5	6
y	2	1,5	1,6	1,4	1

Оценить зависимость себестоимости единицы продукции от величины ее выпуска (считая, что она линейна) и проверить степень соответствия оцененной модели истинной.

1. Итак, зависимая переменная – это себестоимость (y, измеряется в тыс. тг); независимая переменная – это выпуск (x, измеряется в тыс. шт.), а сама модель (регрессии y на x) имеет вид $y = \alpha + \beta x + \varepsilon$, где α и β - неизвестные коэффициенты, а ε - случайный член.

Вычисления оформим в виде таблицы (см. первые 5 столбцов табл. 2.1)

Табл. 2.1

№	$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{xy}$	$\boldsymbol{x^2}$	$\hat{\boldsymbol{y}}$	$\boldsymbol{e}$	$\boldsymbol{e^2}$
1	1	2	2	1	1,968	0,032	0,001
2	3	1,5	4,5	9	1,656	-0,156	0,024
3	5	1,6	8	25	1,344	0,256	0,066
4	5	1,4	7	25	1,344	0,056	0,003
5	6	1	6	36	1,188	-0,188	0,035
∑	**20**	**7,5**	**27,5**	**96**	**7,5**	**0**	**0,13=ESS**

Табл. 2.1 (продолжение)

$x-\bar{x}$	$(x-\bar{x})^2$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$	$y-\bar{y}$	$(y-\bar{y})^2$
-3	9	0,468	0,219	0,5	0,25
-1	1	0,156	0,024	0	0
1	1	-0,156	0,024	0,1	0,01
1	1	-0,156	0,024	-0,1	0,01
2	4	-0,312	0,097	-0,5	0,25
0	**16**	**0**	**0,39=RSS**	**0**	**0,52=TSS**

Найдем сначала неизвестные параметры модели. Для этого применим МНК (см. формулы (*)):

$$\hat{\beta} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} = \frac{5 \cdot 27{,}5 - 20 \cdot 7{,}5}{5 \cdot 96 - 20^2} = -\,0{,}156$$

$$\hat{\alpha} = \frac{\sum y}{n} - \hat{\beta}\frac{\sum x}{n} = \frac{7{,}5}{5} + 0{,}156 \cdot \frac{20}{5} = 2{,}124$$

Т.о., зависимость себестоимости единицы продукции от величины ее выпуска, оцененная по данной выборке, следующая ($\hat{y}$ - прогноз значения зависимой переменной)

$$\hat{y} = 2{,}124 - 0{,}156x. \qquad (2.1)$$

Дадим экономическую интерпретацию этому уравнению. Зависимость между себестоимостью единицы продукции и величиной ее выпуска отрицательная; при увеличении выпуска на одну тысячу штук себестоимость уменьшается (в среднем) на 0,156 тысяч тг, т.е. на 156 тг.

2. Оценим дисперсию ошибок – найдем статистику s^2. Для этого создаем новый столбец – столбец прогнозов зависимой переменной $\hat{y}$, которые вычисляются по формуле (2.1) при различных значениях независимой переменной x:

$$\hat{y}(x_1) = \hat{y}(1) = 2{,}124 - 0{,}156 \cdot 1 = 1{,}968$$

$$\hat{y}(x_2) = \hat{y}(3) = 2{,}124 - 0{,}156 \cdot 3 = 1{,}656$$

$$\hat{y}(x_3) = \hat{y}(5) = 2{,}124 - 0{,}156 \cdot 5 = 1{,}344$$

$$\hat{y}(x_4) = \hat{y}(5) = 1{,}344$$

$$\hat{y}(x_5) = \hat{y}(6) = 2{,}124 - 0{,}156 \cdot 6 = 1{,}188$$

Заметим, что сумма истинных значений зависимой переменной ($\sum y$) равна сумме ее прогнозов ($\sum \hat{y}$). Этот факт имеет место всегда, когда в регрессии есть свободный член.

Далее (см. столбец *e*), вычисляем остатки как разности между истинными значениями зависимой переменной и ее прогнозами:

$$e_t = y_t - \hat{y}_t .$$

Заметим, что в нашем случае (регрессия со свободным членом) сумма остатков – ноль. Остатки e_t, являясь оценками для случайных членов ε_t, принимают как положительные, так и отрицательные значения, но их среднее значение $\bar{e} = \frac{\sum_{t=1}^{n} e_t}{n}$ (а следовательно, и сумма) есть ноль. Если при вычислениях этого не происходит, перепроверьте!

Следующий столбец – квадраты остатков e^2. Заполняем его и находим сумму квадратов остатков $\sum e^2$. Она равна 0,129. Теперь по формуле

$$s^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2}$$

вычисляем статистику s^2 и стандартную ошибку:

$$s^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{0{,}129}{5-2} = 0{,}043\,;\ s = \sqrt{s^2} = \sqrt{0{,}043} = 0{,}207.$$

Найдем стандартные ошибки коэффициентов регрессии по формулам

$$s^2_{\hat{\alpha}} = \frac{s^2 \sum_{t=1}^{n} x_t^2}{n\sum_{t=1}^{n} (x_t - \bar{x})^2}, \qquad s^2_{\hat{\beta}} = \frac{s^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2}. \tag{2.2}$$

Очередной столбец – $x - \bar{x}$ - отклонения переменной *x* от среднего $\bar{x} = \frac{\sum_{t=1}^{n} x_t}{n} = \frac{1+3+5+5+6}{5} = 4$. Для контроля: *сумма отклонений всегда равна нулю!* Следующий столбец – квадраты отклонений $(x - \bar{x})^2$. Их сумма равна 16. По формулам (2.2) имеем:

$$s^2_{\hat{\alpha}} = \frac{s^2 \sum x^2}{n\sum (x - \bar{x})^2} = \frac{0{,}043 \cdot 96}{5 \cdot 16} = 0{,}0516,$$

$$\text{с.о.}(\hat{\alpha}) = \sqrt{s^2_{\hat{\alpha}}} = \sqrt{0{,}0516} = 0{,}227;$$

$$s^2_{\hat{\beta}} = \frac{s^2}{\sum (x - \bar{x})^2} = \frac{0{,}043}{16} = 0{,}0027,$$

$$\text{с.о.}(\hat{\beta}) = \sqrt{s^2_{\hat{\beta}}} = \sqrt{0{,}0027} = 0{,}052.$$

3. Проверим на значимость коэффициенты регрессии.

Найдем *t*-статистики коэффициентов по формулам

$$t_\alpha = \frac{\hat{\alpha}}{\text{с.о.}\left(\hat{\alpha}\right)} = \frac{2{,}124}{0{,}227} = 9{,}357;\ t_\beta = \frac{\hat{\beta}}{\text{с.о.}\left(\hat{\beta}\right)} = \frac{-0{,}156}{0{,}052} = -3.$$

Найдем 5%-е критическое значение статистики Стьюдента с n-2=5-2=3 степенями свободы: $t_{кр}=t_{0,05}(n\text{-}2)=3{,}182$ (используем функцию СТЬЮДОАСПОБР).

Итак, $|\ t_\alpha\ | = 9{,}357 > t_{кр} \Rightarrow \alpha$ значим; $|\ t_\beta\ | = 3 < t_{кр} \Rightarrow \beta$ незначим (на заданном уровне).

4. Для нахождения 95%-х доверительных интервалов воспользуемся необходимыми формулами.

Доверительный интервал для α:

$$(\hat{\alpha} - \text{с.о.}(\hat{\alpha}) \cdot t_{кр};\ \hat{\alpha} + \text{с.о.}(\hat{\alpha}) \cdot t_{кр}) =$$

$$=(2{,}124\text{-}0{,}227\cdot 3{,}182;\ 2{,}124+0{,}227\cdot 3{,}182) = (1{,}402;\ 2{,}846).$$

Видим, что коэффициент α значим (концы его доверительного интервала одного знака).

Доверительный интервал для β:

$$(\hat{\beta} - \text{с.о.}(\hat{\beta}) \cdot t_{кр};\ \hat{\beta} + \text{с.о.}(\hat{\beta}) \cdot t_{кр}) =$$

$$= (-0{,}156\text{-}0{,}052\cdot 3{,}182;\ -0{,}156+0{,}052\cdot 3{,}182)=(-0{,}321;\ 0{,}009).$$

Коэффициент β незначим (концы интервала разных знаков).

5. С помощью соответствующих формул оценим качество аппроксимации данных – найдем коэффициент детерминации. Достаточно применить одну из этих формул, применение другой даст тот же результат (т.к. у нас в регрессии есть свободный член).

Найдем регрессионную сумму квадратов RSS = $\sum_{t=1}^{n}\left(\hat{y}_t - \bar{y}\right)^2$. Для этого организуем столбец отклонений прогнозов от среднего $\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{2+1{,}5+1{,}6+1{,}4+1}{5} = 1{,}5$, озаглавим его $\hat{y} - \bar{y}$ (*сумма отклонений опять-таки ноль*, т.к. выборочное среднее зависимой переменной равно выборочному среднему ее прогнозов: $\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{\sum_{t=1}^{n} \hat{y}_t}{n} = \bar{\hat{y}}$). Вслед за ним – столбец квадратов отклонений - $(\hat{y} - \bar{y})^2$, его сумма, равная 0,39, и есть RSS.

Далее необходимо найти общий разброс зависимой переменной TSS=$\sum_{t=1}^{n}\left(y_t - \bar{y}\right)^2$, для чего мы вводим еще два столбца. Сумма последнего столбца есть TSS, т.е. общий разброс значений зависимой переменной вокруг среднего: TSS=0,52. Выполнено равенство RSS + ESS = TSS (0,39+0,13=0,52).

Применяя первую из формул для нахождения коэффициента детерминации, получаем:

$$R^2 = \frac{\text{RSS}}{\text{TSS}} = \frac{0{,}39}{0{,}52} = 0{,}75.$$

Тот же результат дает применение второй из этих формул:

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{0{,}13}{0{,}52} = 0{,}75.$$

Так как коэффициент детерминации ближе к единице, чем к нулю, делаем вывод, что качество аппроксимации данных lдостаточно хорошее.

6. Проверим на 5%-ю значимость регрессию в целом с помощью *F*-статистики). Имеем:

$$F = (n\text{-}2)\frac{R^2}{1-R^2} = (5\text{-}2)\cdot\frac{0{,}75}{1-0{,}75} = 9;$$

Для нахождения критического значения *F*-статистики используем функцию FРАСПОБР:

$$F_{кр} = F_{0{,}05}(1;\, n\text{-}2) = F_{0{,}05}(1;\, 3) = 10{,}128.$$

Итак, в целом регрессия незначима (так же, как и коэффициент при независимой переменной), т.к. $F < F_{кр}$.

Лабораторная работа № 3 (2 часа)
Классическая модель множественной регрессии. МНК

З а д а н и е. По имеющейся многомерной выборке оценить зависимость переменной $\boldsymbol{y}$ от переменных $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ с помощью МНК и дать экономическую интерпретацию коэффициентам модели.

Общая математическая модель множественной (линейной) регрессии имеет вид

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t \ (t = 1,2,\ldots,n),$$

или в матричном виде

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \qquad (3.1)$$

где $\boldsymbol{y}_{n\times 1} = (y_1,\ldots,y_n)'$- вектор-столбец зависимых (объясняемых) переменных;

$$\boldsymbol{X}_{n\times k} = \begin{pmatrix} x_{11} & \ldots & x_{1k} \\ . & . & . \\ x_{n1} & \ldots & x_{nk} \end{pmatrix}-$$

матрица независимых (объясняющих) переменных, столбцы которой (здесь и далее знак « ′ » означает транспонирование) $\boldsymbol{x_j} = (x_{1j},\ldots, x_{nj})'$ $(j = 1,\ldots,k)$ есть *регрессоры*; $\boldsymbol{\beta}_{k\times 1} = (\beta_1,\ldots,\beta_k)'$ - вектор неизвестных параметров; $\boldsymbol{\varepsilon}_{n\times 1} = (\varepsilon_1,\ldots,\varepsilon_n)'$- вектор случайных членов (ошибок).

Расчетные формулы МНК имеют вид

$$\hat{\beta} = \hat{\beta}_{МНК} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}, \qquad (**)$$

П р и м е р. Оценим зависимость накоплений ($\boldsymbol{y}$) от величины дохода ($\boldsymbol{x}_1$) и стоимости имущества ($\boldsymbol{x}_2$) по данным пяти семей (n = 5, все переменные измеряются в тыс. руб.):

$\boldsymbol{y}$	3	6	5	3	1
$\boldsymbol{x}_1$	40	50	40	30	30
$\boldsymbol{x}_2$	60	40	40	20	90

Итак, модель имеет вид (помимо $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ включаем в модель регрессор-константу; таким образом, в модели *три* регрессора):

$$\boldsymbol{y} = \alpha + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{\varepsilon}$$

и задача состоит в нахождении $\hat{\alpha}$, $\hat{\beta}_1$ и $\hat{\beta}_2$ с помощью МНК (см. формулу (**)).

Шаг 1. Выписываем матрицу регрессоров $\boldsymbol{X}$ (в первом столбце будут одни единицы, т.к. в регрессии есть константа; во втором столбце – значения переменной $\boldsymbol{x}_1$; в третьем – $\boldsymbol{x}_2$; таким образом, искомая матрица будет иметь размерность $n \times k$, или 5×3, т.е. 5 строк и 3 столбца):

$$\boldsymbol{X} = \begin{pmatrix} 1 & 40 & 60 \\ 1 & 50 & 40 \\ 1 & 40 & 40 \\ 1 & 30 & 20 \\ 1 & 30 & 90 \end{pmatrix}.$$

Шаг 2. Транспонируем ее (транспонированная матрица имеет размерность $k \times n$, т.е. 3×5, можно использовать Excel, функция ТРАНСП):

$$\boldsymbol{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 40 & 50 & 40 & 30 & 30 \\ 60 & 40 & 40 & 20 & 90 \end{pmatrix}.$$

Шаг 3. Для получения матрицы плана умножаем $\boldsymbol{X}'$ на $\boldsymbol{X}$. При умножении матрицы размерности $k \times n$ на матрицу размерности $n \times k$ мы получим матрицу $k \times k$, т.е. 3×3 (можно воспользоваться помощью Excel, функция МУМНОЖ):

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 40 & 50 & 40 & 30 & 30 \\ 60 & 40 & 40 & 20 & 90 \end{pmatrix} \cdot \begin{pmatrix} 1 & 40 & 60 \\ 1 & 50 & 40 \\ 1 & 40 & 40 \\ 1 & 30 & 20 \\ 1 & 30 & 90 \end{pmatrix} = \begin{pmatrix} 5 & 190 & 250 \\ 190 & 7500 & 9300 \\ 250 & 9300 & 15300 \end{pmatrix}.$$

Заметим, что полученная матрица квадратная, симметричная и имеет размерность, равную числу регрессоров (у нас три регрессора: регрессор $\boldsymbol{x}_1$, регрессор $\boldsymbol{x}_2$ и регрессор-константа $\boldsymbol{i} = (1,1,1,1,1)'$).

Шаг 4. Находим матрицу, обратную к матрице плана (функция МОБР):

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 7{,}596774 & -0{,}15645 & -0{,}02903 \\ -0{,}15645 & 0{,}003763 & 0{,}000269 \\ -0{,}02903 & 0{,}000269 & 0{,}000376 \end{pmatrix}.$$

Шаг 5. Находим матрицу $\boldsymbol{X}'\boldsymbol{y}$ ($\boldsymbol{X}'_{3\times5} \cdot \boldsymbol{y}_{5\times1} = (\boldsymbol{X}'\boldsymbol{y})_{3\times1}$):

$$\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 40 & 50 & 40 & 30 & 30 \\ 60 & 40 & 40 & 20 & 90 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 6 \\ 5 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 18 \\ 740 \\ 770 \end{pmatrix}.$$

Шаг 6. Находим МНК-оценку, применяя формулу (**):

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = = \begin{pmatrix} 7{,}596774 & -0{,}15645 & -0{,}02903 \\ -0{,}15645 & 0{,}003763 & 0{,}000269 \\ -0{,}02903 & 0{,}000269 & 0{,}000376 \end{pmatrix} \cdot \begin{pmatrix} 18 \\ 740 \\ 770 \end{pmatrix} = \begin{pmatrix} -1{,}3871 \\ 0{,}175806 \\ -0{,}03387 \end{pmatrix}.$$

Итак, $\hat{\beta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} -1{,}3871 \\ 0{,}175806 \\ -0{,}03387 \end{pmatrix}$ и теперь мы можем выписать оцененную модель:

$$\hat{\boldsymbol{y}} = -1{,}387 + 0{,}176\,\boldsymbol{x}_1 - 0{,}034\,\boldsymbol{x}_2,$$

из которой видно, что доход ($\boldsymbol{x}_1$) положительно влияет на накопления, а стоимость имущества ($\boldsymbol{x}_2$) – отрицательно. А именно, при увеличении дохода на одну тысячу рублей накопления будут *увеличиваться* (в среднем) на 0,176 тысяч рублей; при увеличении стоимости имущества на одну тысячу рублей накопления будут *уменьшаться* (в среднем) на 0,034 тысяч рублей.

Лабораторная работа № 4 (6 часов)
Классическая модель множественной регрессии. Основные статистики

З а д а н и е. Проверить соответствие оцененной в предыдущей лабораторной работе модели истинной, т.е.

1. Найти стандартную ошибку и стандартные ошибки коэффициентов регрессии.
2. Проверить на 5%-ю значимость коэффициенты регрессии.
3. Построить 95%-е доверительные интервалы для коэффициентов регрессии и по их виду сделать вывод о значимости коэффициентов.
4. Найти коэффициент детерминации модели и сделать вывод о качестве аппроксимации данных.
5. Проверить на 5%-ю значимость регрессию в целом.

П р и м е р. Напомним оцененную в предыдущей лабораторной работе модель (зависимость накоплений от дохода и стоимости имущества):

$$\hat{\boldsymbol{y}} = -1{,}387 + 0{,}176\,\boldsymbol{x}_1 - 0{,}034\,\boldsymbol{x}_2. \qquad (4.1)$$

Найдем основные статистики модели.

1. Для нахождения оценки для дисперсии ошибок – статистики s^2 – необходимо вычислить прогнозные значения зависимой переменной. Для этого подставляем значения переменных x_1 и x_2 в формулу (4.1):

$$\hat{\boldsymbol{y}}(40;60) = -1{,}387+0{,}176\cdot 40-0{,}034\cdot 60 = 3{,}613.$$

Аналогично находим остальные прогнозы (вычисления сведем в таблицу):

Табл. 4.1

№	$\boldsymbol{y}$	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\hat{\boldsymbol{y}}$	e_t	e_t^2	$y_t - \bar{y}$	$(y_t - \bar{y})^2$
1	3	40	60	3,613	-0,613	0,376	-0,6	0,36
2	6	50	40	6,053	-0,053	0,003	2,4	5,76
3	5	40	40	4,293	0,707	0,500	1,4	1,96
4	3	30	20	3,213	-0,213	0,045	-0,6	0,36
5	1	30	90	0,833	0,167	0,028	-2,6	6,76
∑	**18**			**18**	**0**	**0,95=ESS**	**0**	**15,2=TSS**

Далее находим остатки $e = y - \hat{y}$ и их квадраты. Сумма квадратов остатков 0,95. Делим ее на 2 (объем выборки 5 за вычетом числа регрессоров 3) и получаем статистику $s^2 = \dfrac{\sum_{t=1}^{n} e_t^2}{n-k} = \dfrac{0{,}95}{2} = 0{,}475$. Стандартная ошибка

$$s = \sqrt{0.475} = 0{,}69.$$

Теперь найдем стандартные ошибки коэффициентов по формулам $\text{с.о.}(\hat{\beta}_j) = s\sqrt{(X'X)^{-1}_{jj}}$, т.е. корни из диагональных элементов матрицы $(X'X)^{-1}$ умножаем на стандартную ошибку:

$$\text{с.о.}(\hat{\alpha}) = s\sqrt{(X'X)^{-1}_{11}} = 0{,}689\cdot\sqrt{7{,}597} = 1{,}899;$$

$$\text{с.о.}(\hat{\beta}_1) = s\sqrt{(X'X)^{-1}_{22}} = 0{,}689\cdot\sqrt{0{,}0037} = 0{,}042;$$

$$\text{с.о.}(\hat{\beta}_2) = s\sqrt{(X'X)^{-1}_{33}} = 0{,}689\cdot\sqrt{0{,}00037} = 0{,}013.$$

2. Проверим на 5%-ю значимость коэффициенты, вычислив их t-статистики t_α и $t_{\beta_j} = \dfrac{\hat{\beta}_j}{\text{с.о.}\left(\hat{\beta}_j\right)}$ $(j = 1,2)$:

$$t_\alpha = \frac{-1{,}387}{1{,}899} = -0{,}730;\ t_{\beta_1} = \frac{0{,}176}{0{,}042} = 4{,}190;\ t_{\beta_2} = \frac{-0{,}034}{0{,}013} = -2{,}615;$$

$$t_{кр} = t_{0,05}(n\text{-}k) = t_{0,05}(5\text{-}3) = t_{0,05}(2) = 4{,}303.$$

$$|t_\alpha| = 0{,}730 < t_{кр};\ |t_{\beta_1}| = 4{,}190 < t_{кр};\ |t_{\beta_2}| = 2{,}615 < t_{кр}.$$

Вывод: все коэффициенты незначимы. Заметим, что это может являться следствием недостаточного объема выборки (мы рассматриваем модельную ситуацию).

3. Построим 95%-е доверительные интервалы для коэффициентов регрессии. Имеем:

$$(\hat{\alpha} - \text{с.о.}(\hat{\alpha})\cdot t_{кр};\ \hat{\alpha} + \text{с.о.}(\hat{\alpha})\cdot t_{кр})$$

$$(\hat{\beta}_j - \text{с.о.}(\hat{\beta}_j)\cdot t_{кр};\ \hat{\beta}_j + \text{с.о.}(\hat{\beta}_j)\cdot t_{кр})\ (j = 1,2),$$

т.е. для α: (-1,387-1,899·4,303;-1,387+1,899·4,303)=(-9,558;6,784);

β_1: (0,176-0,042·4,303;0,176+0,042·4,303)=(-0,005; 0,357);

β_2:(-0,034-0,013·4,303;-0,034+0,013·4,303)=(-0,090; 0,022).

Все интервалы имеют концы разных знаков (содержат ноль), что подтверждает 5%-ю незначимость коэффициентов, выявленную в предыдущем пункте.

4. Для нахождения коэффициента детерминации модели нам остается найти TSS (ESS уже найдено). Имеем ($\bar{y} = \sum y/5 = 18/5 = 3{,}6$; далее см. табл. 4.1): $R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{0{,}95}{15{,}2} = 0{,}9375$. Итак, качество аппроксимации данных достаточно хорошее, т.к. коэффициент детерминации близок к единице.

5. Наконец, проверим на значимость регрессию в целом. Вычисляем F-статистику

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0{,}9375/(3-1)}{(1-0{,}9375)/(5-3)} = 15.$$

Сравниваем ее с критической $F_{кр} = F_{0,05}(k\text{-}1;n\text{-}k) = F_{0,05}(3\text{-}1;\ 5\text{-}3) = F_{0,05}(2,2) = 19$ и делаем вывод, что регрессия в целом незначима.

Лабораторная работа № 5 (2 часа)
Нелинейная регрессия. Нелинейность по переменным

З а д а н и е. По выборочным данным оценить две модели: линейную модель $y = \alpha + \beta_1 x + \varepsilon$ и нелинейную по независимым переменным модель $y = \alpha + \beta_1 x + \beta_2 x^2 + \varepsilon$. Сравнить качество аппроксимации данных.

П р и м е р. Выборочные данные следующие:

y	6	16	35	55	86
x	1	2	3	4	5

Оценим сначала линейную модель $y = \alpha + \beta_1 x + \varepsilon$ (см. лабораторную работу № 1).

Имеем таблицу:

№	y_t	x_t	x_t^2	$x_t y_t$	$\hat{y}_t$	e_t	e_t^2	$y_t - \bar{y}$	$(y_t - \bar{y})^2$
1	6	1	1	6	-0,2	6,2	38,44	-33,6	1128,96
2	16	2	4	32	19,7	-3,7	13,69	-23,6	556,96
3	35	3	9	105	39,6	-4,6	21,16	-4,6	21,16
4	55	4	16	220	59,5	-4,5	20,25	15,4	237,16
5	86	5	25	430	79,4	6,6	43,56	46,4	2152,96
Σ	**198**	**15**	**55**	**793**	**198**	**0**	**137,1**	**0**	**4097,2**
							ESS		**TSS**

$$\hat{\beta} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} = 19{,}9;\ \hat{\alpha} = \frac{\sum y}{n} - \hat{\beta}\frac{\sum x}{n} = -20{,}1$$

Оцененная модель:

$$\hat{y} = -20{,}1 + 19{,}9x.$$

Для оценки качества аппроксимации данных вычислим коэффициент детерминации модели:

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{137{,}1}{4097{,}2} = 0{,}9665.$$

Теперь оценим нелинейную модель $y = \alpha + \beta_1 x + \beta_2 x^2 + \varepsilon$, приведя ее предварительно к линейному виду. Заметим, что нелинейные по независимым переменным модели линеаризуются с помощью соответствующей замены переменной. В данном случае необходимо сделать замену $z = x^2$, после чего модель станет линейной, а регрессия – множественной: $y = \alpha + \beta_1 x + \beta_2 z + \varepsilon$.

Теперь необходимо ввести новый регрессор $z = x^2$, после чего вышеописанными методами (см. лабораторную работу № 4) оценить преобразованную модель.

Имеем:

№	y	x	$z = x^2$	$\hat{y}$	e	e^2	$y - \bar{y}$	$(y - \bar{y})^2$
1	6	1	1	5,942857	0,057143	0,003265	-33,6	1128,96
2	16	2	4	16,62857	-0,62857	0,395102	-23,6	556,96
3	35	3	9	33,45714	1,542857	2,380408	-4,6	21,16
4	55	4	16	56,42857	-1,42857	2,040816	15,4	237,16
5	86	5	25	85,54286	0,457143	0,20898	46,4	2152,96
Σ	**198**	**15**	**55**	**198**	**0**	**5,028571**	**0**	**4097,2**
						ESS		TSS

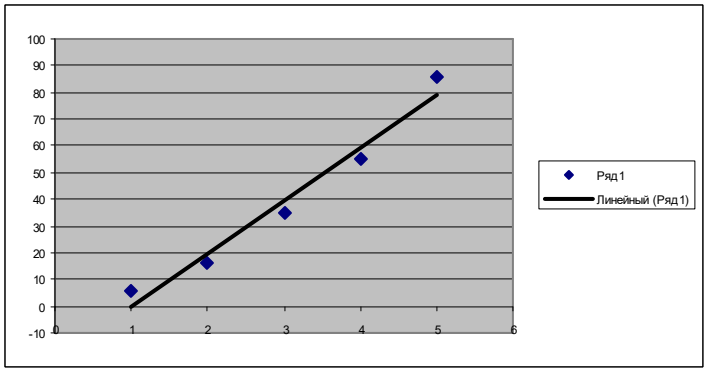
Далее находим коэффициент детерминации:

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{5{,}028571}{4097{,}2} = 0{,}9988.$$

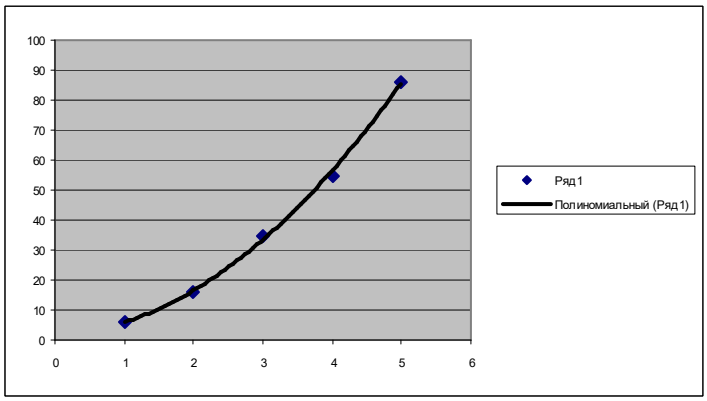
Из сравнения коэффициентов детерминации двух оцененных моделей видим, что качество аппроксимации данных нелинейной функцией лучше, чем линейной, т.е. данную зависимость лучше приближать нелинейной функцией (в нашем случае это полином второй степени – парабола).

Можно также вывести на экран графики полученных зависимостей:

Линейная модель:



Нелинейная модель:



Из сравнения графиков также видно, что парабола лучше приближает зависимость, представленную нашей выборкой.

Лабораторная работа № 6 (4 часа)
Нелинейная регрессия. Нелинейность по параметрам. Производственная функция

З а д а н и е. Исследовать производственную функцию.

1. По выборочным данным оценить модель $\boldsymbol{Q} = A\boldsymbol{K}^{\alpha}\boldsymbol{L}^{\beta}\boldsymbol{v}$ (производственную функцию Кобба-Дугласа), где $\boldsymbol{Q}$ – выпуск продукции, $\boldsymbol{K}$ – затраты капитала, $\boldsymbol{L}$ – затраты труда, $\boldsymbol{v}$ – случайный член.

2.Выписать коэффициенты эластичности выпуска по капиталу и по труду, интерпретировать их.

3. Проверить на 5%-ю значимость коэффициенты регрессии.

4. Найти коэффициент детерминации модели.

5. Проверить на 5%-ю значимость регрессию в целом.

П р и м е р. Выборочные данные по областям Казахстана (2001 г.) имеют вид (***Q*** – объем выполненных строительных работ (млн тг), ***K*** – инвестиции в жилищное строительство (млн тг), ***L*** – количество занятых в жилищном строительстве человек (тыс.):

№	Области	***Q***	***K***	***L***
1	Акмолинская	1989	3806	7,6
2	Актюбинская	6922	49 716	8
3	Алматинская	11957	13 687	27,4
4	Атырауская	44019	172 228	20,1
5	Восточно-Казахстанская	14616	24 742	20,1
6	Жамбыльская	3688	2 971	15,2
7	Западно-Казахстанская	46747	111 341	22,1
8	Карагандинская	9223	34 610	17,9
9	Костанайская	3945	11 501	11
10	Кызылординская	4984	18 407	9,2
11	Мангистауская	14507	37 507	10,1
12	Павлодарская	4996	18 877	15,2
13	Северо-Казахстанская	1833	3 648	5,2
14	Южно-Казахстанская	7972	20378	23,2
15	г. Астана	37135	54385	18,2
16	г. Алматы	39157	27111	33,7

1. Модель $\boldsymbol{Q} = A\boldsymbol{K}^{\alpha}\boldsymbol{L}^{\beta}\boldsymbol{v}$ относится к классу нелинейных по параметрам, но внутренне линейных моделей. Для линеаризации таких моделей обычно применяется соответствующее элементарное преобразование, затем – замена переменных.

Прологарифмируем модель: $\ln \boldsymbol{Q} = \ln A + \alpha \ln \boldsymbol{K} + \beta \ln \boldsymbol{L} + \ln \boldsymbol{v}$. Теперь введем новые обозначения: $\boldsymbol{y} = \ln \boldsymbol{Q}$, $\boldsymbol{x}_1 = \ln \boldsymbol{K}$, $\boldsymbol{x}_2 = \ln \boldsymbol{L}$, $a = \ln A$, $\boldsymbol{\varepsilon} = \ln \boldsymbol{v}$. В этих обозначениях модель принимает вид:

$$\boldsymbol{y} = a + \alpha \boldsymbol{x}_1 + \beta \boldsymbol{x}_2 + \boldsymbol{\varepsilon},$$

т.е. является линейной.

Выборочные данные после логарифмирования приобретают вид:

y	x_1	x_2
7,595387	8,244334	2,028148
8,84246	10,81408	2,079442
9,389072	9,524202	3,310543
10,69238	12,05657	3,00072
9,589872	10,11626	3,00072
8,21284	7,996654	2,721295
10,75251	11,62035	3,095578
9,129456	10,4519	2,884801
8,280204	9,350189	2,397895
8,513988	9,820486	2,219203
9,582387	10,53228	2,312535
8,516393	9,8457	2,721295
7,513709	8,201934	1,648659
8,983691	9,922211	3,144152
10,52232	10,90384	2,901422
10,57533	10,20769	3,517498

Теперь для оценки модели применяем методы, описанные в лабораторных работах №№ 4 и 5, и получаем уравнение:

$$\hat{y} = 0{,}7536 + 0{,}6003\, x_1 + 0{,}9031\, x_2$$

или

$$\ln \hat{Q} = 2{,}7536 + 0{,}6003\ \ln K + 0{,}9031\ \ln L,$$

т.е.

$$\hat{Q} = 2{,}125 K^{0{,}6003} L^{0{,}9031}$$

2. Коэффициент эластичности объема строительных работ (Q) по инвестициям (K) есть коэффициент при $\ln K$, т.е. 0,6003, что означает: при увеличении инвестиций на 1% объем работ будет увеличиваться в среднем на 0,6003%.

Коэффициент эластичности объема строительных работ (Q) по количеству занятых (L) есть коэффициент при $\ln L$, т.е. 0,9031, что означает: при увеличении количества занятых на 1% объем работ будет увеличиваться в среднем на 0,9031%.

3. Так же, как в лабораторной работе №4 (п. 2), проверяем на значимость коэффициенты регрессии ($t_{кр} = t_{0,05}(16\text{-}3) = t_{0,05}(13) = 2{,}160$):

$t_a = 0{,}815634$, $|t_a| < t_{кр} \Rightarrow a$ незначим;

$t_\alpha = 5{,}925802$ $|t_\alpha| > t_{кр} \Rightarrow \alpha$ значим;

$t_\beta = 4{,}027043$ $|t_\beta| > t_{кр} \Rightarrow \beta$ значим.

4. Коэффициент детерминации модели вычисляем по формулам, приведенным в лабораторной работе № 4 (п. 4):

$$R^2 = 0{,}8742,$$

что свидетельствует о достаточно хорошем качестве аппроксимации данных.

5. Наконец (см. лаб. лаб. № 4, п.5), проверяем на значимость регрессию в целом ($F_{кр} = F_{0,05}(3\text{-}1;16\text{-}3) = F_{0,05}(2;13) = 3{,}806$):

$$F = 45{,}17785 > F_{кр} \Rightarrow \text{регрессия в целом значима.}$$

Исследование модели закончено.

Лабораторная работа № 7 (2 часа)
Фиктивные переменные. Исследование сезонности

З а д а н и е. Исследовать объем продаж данного товара на наличие сезонности (уровень значимости теста 5%).

П р и м е р. Выборочные (помесячные) данные по объему продаж данного товара имеют вид:

Месяц	Объем продаж (млн тг)
Январь	2
Февраль	2
Март	3
Апрель	4
Май	3
Июнь	4
Июль	5
Август	5
Сентябрь	3
Октябрь	4
Ноябрь	3
Декабрь	2

Пусть y_t – объем продаж товара в месяц t и есть основания считать, что объем продаж зависит от времени года. Для выявления сезонности введем 3 фиктивные (бинарные) переменные

$$d_{t1} = \begin{cases} 1, \text{ если } t - \text{зимний месяц}, \\ 0, \text{ если нет}; \end{cases} \quad d_{t2} = \begin{cases} 1, \text{ если } t - \text{весенний месяц}, \\ 0, \text{ если нет}; \end{cases}$$

$$d_{t3} = \begin{cases} 1, \text{ если } t - \text{летний месяц}, \\ 0, \text{ если нет}. \end{cases}$$

Получаем модель:

$$y_t = \alpha + \beta_1 d_{t1} + \beta_2 d_{t2} + \beta_3 d_{t3} + \varepsilon_t \quad (t = 1,2,\ldots,12).$$

З а м е ч а н и е. Мы не вводим фиктивную переменную d_{t4}, соответствующую осенним месяцам, т.к. в этом случае при любом t=1,…,12 будем иметь

$$d_{t1} + d_{t2} + d_{t3} 1 d_{t1} + d_{t4} = 1,$$

что означает линейную зависимость регрессоров и, как следствие, неприменимость МНК.

Таким образом, среднемесячное потребление данного продукта есть α (осенью); $\alpha + \beta_1$ (зимой); $\alpha + \beta_2$ (весной) и $\alpha + \beta_3$ (летом). Коэффициенты β_1, β_2 и β_3 показывают среднесезонные колебания потребления по отношению к осенним месяцам.

Теперь нам необходимо оценить модель и протестировать гипотезу $\beta_1=\beta_2=\beta_3=0$, т.е. проверить предположение об отсутствии сезонности в объеме продаж данного товара. Проверка этой гипотезы есть не что иное, как проверка на значимость регрессии в целом.

Итак, после введения фиктивных переменных матрица регрессоров $\boldsymbol{X}$ примет вид

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

где первый столбец – это регрессор-константа, второй – регрессор $\boldsymbol{d}_1$, соответствующий зимним месяцам, третий – $\boldsymbol{d}_2$, соответствующий весенним месяцам и четвертый – $\boldsymbol{d}_3$, соответствующий летним месяцам.

Далее с помощью МНК (см. лаб. раб. №3) оцениваем модель:

$$\hat{\boldsymbol{y}} = 3{,}333 - 1{,}333\boldsymbol{d}_1 + 0\boldsymbol{d}_2 + 1{,}333\boldsymbol{d}_3 \quad (t = 1,2,\ldots,12).$$

Теперь проверим на значимость регрессию в целом. Для этого вычисляем F-статистику (см. лаб. раб. №5, п. 5):

$$F = 10{,}667,$$

находим $F_{кр} = F_{0,05}(4\text{-}1;\ 12\text{-}4) = F_{0,05}(3;\ 8) = 5{,}318$ и, т.к. $F > F_{кр}$, делаем вывод о наличии сезонности в продажах данного товара.

Лабораторная работа № 8 (2 часа)
Коэффициент частной корреляции

З а д а н и е. По выборке ($\boldsymbol{y}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$) найти выборочный коэффициент частной корреляции между зависимой переменной $\boldsymbol{y}$ и независимой переменной $\boldsymbol{x}_1$, исключив влияние переменной $\boldsymbol{x}_2$.

П р и м е р. Выборка наблюдений по пяти семьям имеет вид (в условных денежных единицах):

Семья № t	y_t	x_{t1}	x_{t2}
1	3	40	60
2	6	55	36
3	5	45	36
4	3,5	30	15
5	1,5	30	90

где y_t – накопления семьи № t, x_{t1} – ее доход, x_{t2}- стоимость имущества.

Необходимо оценить модель регрессии $\boldsymbol{y}$ на $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ и константу и найти выборочный коэффициент частной корреляции между накоплениями $\boldsymbol{y}$ и доходом $\boldsymbol{x}_1$ при исключении влияния стоимости имущества $\boldsymbol{x}_2$, т.е. $r(y, x_1/ x_2)$.

Процедура заключается в следующем.

1) Оценивается регрессия $\boldsymbol{y}$ на $\boldsymbol{x}_2$ и константу $\hat{\boldsymbol{y}} = \hat{\alpha} + \hat{\beta}\boldsymbol{x}_2$ (см. лаб. раб. №2) и находится вектор остатков $\boldsymbol{e}_1$:

$$\hat{\boldsymbol{y}} = 5{,}779 - 0{,}042\boldsymbol{x}_2,$$

$$\boldsymbol{e}_1 = \begin{pmatrix} -0{,}27385 \\ 1{,}72396 \\ 0{,}72396 \\ -1{,}65296 \\ -0{,}52111 \end{pmatrix}$$

(заметим, что $\sum_{t=1}^{5} e_{t1} = 0$, т.к. в регрессии есть константа).

2) Оценивается регрессия $\boldsymbol{x}_1$ на $\boldsymbol{x}_2$ и константу $\hat{\boldsymbol{y}} = \hat{\gamma} + \hat{\delta}\boldsymbol{x}_2$, находится вектор остатков $\boldsymbol{e}_2$:

$$\hat{\boldsymbol{x}}_1 = 44{,}764 - 0{,}101\boldsymbol{x}_2,$$

$$\boldsymbol{e}_2 = \begin{pmatrix} 1{,}26645 \\ 13{,}85417 \\ 3{,}85417 \\ -13{,}2566 \\ -5{,}7182 \end{pmatrix}, \quad \sum_{t=1}^{5} e_{t2} = 0$$

Т.к. остатки и регрессоры линейно независимы, для исключения влияния $\boldsymbol{x}_2$ вычисляется выборочный коэффициент корреляции между остатками $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$ (они не коррелированы с $\boldsymbol{x}_2$) r_{e_1,e_2} (можно использовать функцию Excel КОРРЕЛ).

Итак, выборочный коэффициент частной корреляции между y и x_1 при исключении влияния x_2 есть

$$r(\boldsymbol{y}, \boldsymbol{x}_1/ \boldsymbol{x}_2) = r_{e_1,e_2} = 0{,}978,$$

а его близость к единице говорит о достаточно хорошей стохастически-линейной связи между $\boldsymbol{y}$ и $\boldsymbol{x}_1$ при исключении влияния $\boldsymbol{x}_2$.

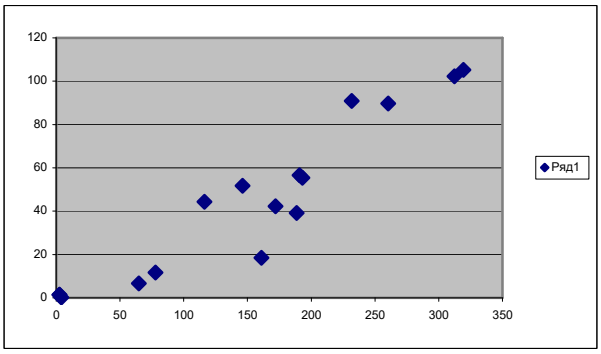
Лабораторная работа № 9 (2 часа)
Тест на гетероскедастичность Голдфелда-Куандта

З а д а н и е. По имеющимся выборочным данным определить, присутствует ли в модели гетероскедастичность с помощью теста Голдфелда-Куандта (уровень значимости 5%).

П р и м е р. Имеются выборочные данные (по Казахстану) по поголовью коров ($\boldsymbol{x}$) и производству молока ($\boldsymbol{y}$):

Коровы (тыс. гол.)	Молоко (тыс. т.)
190,9	56,6
171,9	42,2
319,4	105,2
64,9	6,7
312,4	102,3
116,3	44,3
161	18,4
188,6	39,1
231,7	90,9
77,8	11,7
4,1	0,2
146,1	51,7
193	55,4
260,3	89,7
2,5	1,5
2,6	1,3

Проверим, есть ли в модели гетероскедастичность с помощью теста Голдфелда-Куандта. Сначала выведем диаграмму рассеяния выборки:



По виду диаграммы рассеяния можно сделать вывод как в пользу наличия гетероскедастичности (неодинакового разброса данных), так и в пользу ее отсутствия. Применим тест.

Процедура теста заключается в следующем.

1. Данные упорядочиваются по возрастанию регрессора ***x***, относительно которого есть подозрения, что именно он вызывает гетероскедастичность:

x	***y***
2,5	1,5
2,6	1,3
4,1	0,2
64,9	6,7
77,8	11,7
116,3	44,3
146,1	51,7
161	18,4
171,9	42,2
188,6	39,1
190,9	56,6
193	55,4
231,7	90,9
260,3	89,7
312,4	102,3
319,4	105,2

2. Из общего количества n = 16 наблюдений исключаются d средних ($d \approx n/4 = 4$).

3. Применяется МНК (см. лаб. раб. № 2) отдельно к оставшимся первым $(n-d)/2=6$ и последним $(n-d)/2=6$ наблюдениям, находятся суммы квадратов остатков ESS_1 и ESS_2 (соответственно): ESS_1 = 352,6649; ESS_2 = 312,4959.

4. Вычисляется статистика $F = ESS_2/ESS_1$ (или ESS_1/ESS_2, числитель должен быть больше знаменателя). При справедливости гипотезы об

отсутствии гетероскедастичности эта статистика имеет распределение Фишера (k – число регрессоров) $F((n\text{-}d)/2\text{-}k, (n\text{-}d)/2\text{-}k) = F(4, 4)$. Если $F > F_{кр} = F_{0,05}(4;4)$, то нулевая гипотеза отвергается и считается, что гетероскедастичность имеет место.

Однако в нашем случае $F = ESS_1 / ESS_2 = 352,6649 / 312,4959 = 1,042$, $F_{кр} = 6,3882$; $F < F_{кр} \Rightarrow$ гетероскедастичности нет.

З а м е ч а н и е. Этот тест работает в случае, когда гетероскедстичность зависит от регрессора и в случае, когда дисперсия ошибок принимает 2 значения. Возможно, в нашей выборке присутствует другой вид гетероскедастичности, который можно выявить с помощью других тестов.

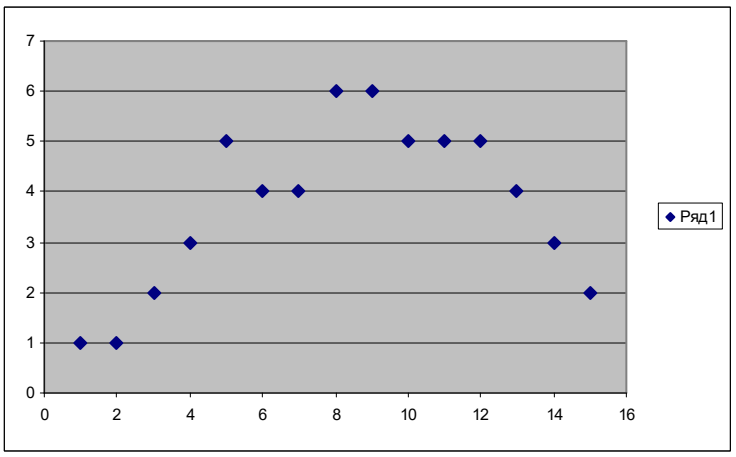
Лабораторная работа № 10 (2 часа)
Тест на автокорреляцию Дарбина-Уотсона

З а д а н и е. С помощью теста Дарбина-Уотсона определить, есть ли в модели, представленной выборочными данными, автокорреляция ошибок.

П р и м е р. Имеется выборка

y	***x***
1	1
1	2
2	3
3	4
5	5
4	6
4	7
6	8
6	9
5	10
5	11
5	12
4	13
3	14
2	15

Диаграмма рассеяния выборки имеет вид:



Мы предполагаем, что здесь имеет место авторегрессионный процесс I порядка. Для выявления этого факта применим тест Дарбина-Уотсона, который основан на вычислении статистики *DW*. Имеем (e_t – МНК-остатки регрессии $\boldsymbol{y}$ на $\boldsymbol{x}$ с константой):

y_t	x_t	$x_t y_t$	x_t^2	$\hat{y}_t$	e_t	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
1	1	1	1	2,7083333	-1,7083333	2,9184028		
1	2	2	4	2,8547619	-1,8547619	3,4401417	-0,1464286	0,0214413
2	3	6	9	3,0011905	-1,0011905	1,0023824	0,8535714	0,7285842
3	4	12	16	3,147619	-0,147619	0,0217914	0,8535714	0,7285842
5	5	25	25	3,2940476	1,7059524	2,9102735	1,8535714	3,435727
4	6	24	36	3,4404762	0,5595238	0,3130669	-1,1464286	1,3142985
4	7	28	49	3,5869048	0,4130952	0,1706477	-0,1464286	0,0214413
6	8	48	64	3,7333333	2,2666667	5,1377778	1,8535714	3,435727
6	9	54	81	3,8797619	2,1202381	4,4954096	-0,1464286	0,0214413
5	10	50	100	4,0261905	0,9738095	0,948305	-1,1464286	1,3142985
5	11	55	121	4,172619	0,827381	0,6845592	-0,1464286	0,0214413
5	12	60	144	4,3190476	0,6809524	0,4636961	-0,1464286	0,0214413
4	13	52	169	4,4654762	-0,4654762	0,2166681	-1,1464286	1,3142985
3	14	42	196	4,6119048	-1,6119048	2,598237	-1,1464286	1,3142985
2	15	30	225	4,7583333	-2,7583333	7,6084028	-1,1464286	1,3142985
56	120	489	1240	56	0	32,929762	-1,05	15,007321

Оцененная модель: $\hat{\boldsymbol{y}} = 2{,}5619 + 0{,}1464\boldsymbol{x}$.

$$DW = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = 0{,}4557.$$

Существуют таблицы для определения нижней d_l и верхней d_u критических границ статистики DW, при сравнении с которыми вычисленной статистики делаются соответствующие статистические выводы, а именно:

- если $DW<d_l$, то принимается гипотеза о положительной автокорреляции;
- если $d_l<DW<d_u$, то тест не работает (зона неопределенности);
- если $d_u<DW<4-d_u$, то принимается гипотеза об отсутствии автокорреляции;
- если $4-d_u<DW<4-d_l$, то тест не работает (зона неопределенности);
- если $4-d_l<DW<4$, то принимается гипотеза об отрицательной автокорреляции.

В нашем случае d_l = 1,08; d_u = 1,36 и $DW<d_l$, т.е. имеет место положительная автокорреляция.

ЛИТЕРАТУРА

1. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. М.: 2000
2. Катышев П.К., Пересецкий А.А. Сборник задач к начальному курсу эконометрики. М.: 2003
3. Елисеев И.И. Эконометрика. М.: 2005
4. Бородич С.А. Эконометрика. Мн.: 2004
5. Колемаев В.А. Эконометрика. М.: 2004
6. Ковалева И.М. Введение в регрессионный анализ. Алматы: 2007