Практическая работа 1

Тема: Анализ однородности статистических рядов

Содержание работы:

1. С помощью критерия Фишера определить однородность рядов расхода воды р.Черный Иртыш у с.Буран за период 1938-1950 и 1977-1989 гг. Определить однородность расхода воды р.Черный Иртыш названных рядов с помощью критерия Стьюдента и по методу наименьшей существенной разности (НСР) при 5% и 1% уровня значимости.

Ход работы:

1. По имеющимся данным о стоке р.Черный Иртыш вычислить среднее квадратическое отклонение годовых расходов воды.

Таблица 1.1

Вычисление среднего квадратического отклонения годовых расходов воды р.Черный Иртыш у с.Буран 1938-1950 и 1977-1989 гг.

n	Год	Q_i	Q_i - M_1	$(Q_i - M_1)^2$	n	Год	Q_i	Q_i - M_2	$(Q_i - M_2)^2$
1	1938	252			1	1977			
2	1939	281			2	1978			
3	1940	327			3	1979			
4	1941	410			4	1980			
5	1942	425			5	1981			
6	1943	307			6	1982			
7	1944	267			7	1983			
8	1945	218			8	1984			
9	1946	458			9	1985			
10	1947	372			10	1986			
11	1948	239			11	1987			
12	1949	274			12	1988			
13	1950	383			13	1989			
				Σ=					Σ=

$$n = 13$$

Вычислить среднее арифметическое значение (М) рядов:

$$M_1 = \frac{\sum_1^{13} Q_{i1}}{n}$$

$$M_2 = \frac{\sum_1^{13} Q_{i2}}{n}$$

Вычислить среднее квадратическое отклонение (σ) рядов:

$$\sigma_1 = \sqrt{\frac{\sum(Q_i - M_1)^2}{n-1}}$$

$$\sigma_2 = \sqrt{\frac{\sum(Q_i - M_2)^2}{n-1}}$$

2. Рассчитать критерий Фишера (F) и определить однородность рядов, если табличное значение критерия Фишера составляет 2,65 (величина степеней свободы рассчитывается по формуле: v = n-1)

$F_ф = (\sigma_1^2) / (\sigma_2^2)$;

При $F_ф > F_т$ – ряд неоднородный, при $F_ф < F_т$ – ряд однородный.

3. Рассчитать критерий Стьюдента ($t_ф$) и определить однородность рядов $t_ф = (M_1 - M_2) / m_\Sigma$, где m_Σ – суммарная ошибка

$m_\Sigma = \sqrt{m_1^2 + m_2^2}$;

$m_1 = \frac{\sigma_1}{\sqrt{n}}$ $m_1 = \frac{\sigma_2}{\sqrt{n}}$

Если $t_ф > t_т$, то ряд неоднородный, $t_ф < t_т$ – ряд однородный

$t_т = 2{,}18$ (5%), $t_т = 3{,}05$ (1%)

4.Рассчитать показатель наименьшей существенной разности, определить однородность рядов.
НСР = $t_т * m_\Sigma$; $t_т$ – табличное значение критерия Стьюдента при выбранном уровне вероятности.

Если $M_1 - M_2 <$ НСР, то ряд однородный,

$M_1 - M_2 >$ НСР – ряд неоднородный.

ВЫВОД.

Практическая работа 2

Тема: Однофакторный дисперсионный анализ

Содержание работы:

По данным урожайности ячменя при внесении в почву, при прочих равных условиях разных доз торфа (100, 150, 300 т абсолютно сухого вещества на гектар) при внесении его на фоне минеральных, органических удобрений и доломитовой муки. Исходная почва – дерново-подзолистая глееватая связносупесчаная осушенная.

1. Построить однофакторный дисперсионный комплекс.
2. Вычислить девиаты (суммы квадратических отклонений от среднего) и дисперсии результатов опыта по общему варьированию, по вариантам опыта, по повторностям и по остаточному варьированию.
3. С помощью критерия Фишера определить достоверность влияния доз торфа, внесенного в почву на урожай ячменя.
4. Полученные результаты проверить по методу наименьшей существенной разности.
5. Вычислить показатель точности опыта и оценить его достоверность.

Таблица 1 Результат опытов по изучению влияния на урожай дополнительных доз торфа.

Варианты опыта	Урожай ячменя по повторностям			
	1	2	3	4
Контрольный фон	10	11	12	10
Фон+100 т/га	15	17	18	16
Фон+150 т/га	20	28	28	29
Фон+300 т/га	25	28	29	29

После получения сведений об урожайности ячменя в названных условиях составляется таблица дисперсионного комплекса (табл. 2.), куда заносится исходная информация по группам влияющего фактора (вариантам опыта) и некоторые результаты расчетов (для удобства сделано округление по урожайности до целых чисел). Вначале производим расчет данных по вариантам опыта (строкам).

Результаты разносим по столбцам. Суммарный урожай ячменя по повторностям Σx_i и по каждому варианту опыта вносим в столбец 6 в числителе. Аналогично поступаем с квадратами этих показателей Σx_i^2. Затем в столбце 7 приводим квадраты суммарного урожая ячменя по повторностям $(\Sigma x_i)^2$. И, наконец, вычисляем среднее арифметическое M_i по каждому варианту опыта, заносим в столбец 8; вычисляем общее среднее $M_{общ}$.

После получения данных по вариантам опыта производим расчет необходимых показателей по повторностям (x_k). Сначала суммируем данные урожайности ячменя и приводим в строке под чертой Σx_k. Суммы сумм урожайности ячменя по вариантам опыта и повторностям должны совпасть и дать сумму всех вариант ($\Sigma\Sigma x_{i,k}$). Аналогично суммируем квадраты этих

показателей по повторностям (Σx_k^2). Суммы сумм квадратов по вариантам и повторностям опыта должны совпасть и дать сумму квадратов всех вариант ($\Sigma x_i^2 = \Sigma x_k^2$). Ниже вписываем результаты возведения в квадрат сумм вариант по каждой повторности $(\Sigma x_k)^2$ и суммируем их: $\Sigma(\Sigma x_k)^2$. Вычисляем средние арифметические по каждой повторности опыта M_k. Общее среднее арифметическое всех вариант опыта $M_{общ} = (\Sigma x_{i,k})/N$.

Таблица 2 **Дисперсионный комплекс для анализа влияния внесения доз торфа на урожай ячменя**

Варианты опыта (фактор)		Урожай ячменя по повторностям, ц/га*				По повторностям (признакам) (i)		
						$\frac{\sum x_i}{\sum(x_i^2)}$	$(\sum x_i)^2$	M_i
Контроль (фон)		$\frac{10}{100}$						
Фон+100 т/га торфа		$\frac{15}{}$						
Фон+150 т/га торфа		$\frac{20}{}$						
Фон+300 т/га торфа		$\frac{25}{}$						
По факторам	$\sum x_k$					$\sum\sum x_{i,k}$	$\sum(\sum x_i)^2$	
	$\sum(x_k^2)$					$\sum(\sum x_{i,k}^2)$		
	$(\sum x_k)^2$							
	M_k					$\sum(\sum x_k)^2 =$		$M_{общ=}$

Примечание: * В числителе – опытные данные, в знаменателе – квадраты этих показателей.

Следующий этап работы – нахождение сумм квадратов отклонений, т. е. расчленение общего варьирования признака на составные части исходя из равенства:

$$\Theta = \Theta_1 + \Theta_2 + \Theta_3,$$

где Θ – сумма квадратов отклонений по общему варьированию данных, Θ_1 – по группам фактора (варианты опыта), Θ_2 – по повторностям опыта, Θ_3 – по остаточному варьированию, вызванному неучтенными факторами.

Общая сумма квадратов отклонений вычисляется следующим образом:

$$\Theta = \Sigma(\Sigma x^2_{i,k}) - (\Sigma\Sigma x_{i,k})^2 / N$$

Затем находим сумму квадратов отклонений по группам фактора (варианты опыта) по формуле:

$$\Theta_1 = [\Sigma(\Sigma x_i)^2 - (\Sigma\Sigma x_{i,k})^2/k] / i,$$

где k – число групп фактора, т. е. 4; i – число повторностей, т. е. 4. В данном случае должно выдержаться равенство $N = ki$.

Сумму квадратов отклонений по повторностям опыта находим по формуле

$$\Theta_2 = [\Sigma(\Sigma x_k)^2 - (\Sigma\Sigma x_{i,k})^2/i] / k,$$

где i – число повторностей, т. е. 4; k — число слагаемых в каждой сумме Σx_k

Таблица 3 **Результаты однофакторного дисперсионного анализа**

Варьирование данных	Сумма квадратов отклонений Θ	Степень свободы ν	Дисперсия $\sigma^2 = \Theta/\nu$	Критерий Фишера	
				$F_ф$	$F_т$
Общее по опыту				–	–
По вариантам опыта					8,81
По повторностям					8,81
Случайное (остаточное)				–	–

Сумма квадратов отклонений по остаточному варьированию определяется из равенства

$$\Theta_3 = \Theta - \Theta_1 - \Theta_2.$$

Проводим дисперсионный анализ данных урожая ячменя (табл. 3). Вносим в таблицу рассчитанные суммы квадратов отклонений (Θ, Θ_1, Θ_2, Θ_3). Число степеней свободы получаем следующим образом: по общей сумме квадратов отклонений $\nu = N - 1$ по вариантам опыта $\nu_1 = n_1 - 1$ по повторностям $\nu_2 = n_2 - 1$ по остаточной сумме $\nu_3 = \nu - \nu_1 - \nu_2$ Дисперсия определяется путем деления сумм квадратов отклонений (Θ, Θ_1, Θ_2, Θ_3) на соответствующие им числа степеней свободы (ν, ν_1, ν_2, ν_3), что можно выразить в общем виде формулой $\sigma^2 = \Theta/\nu$.

Оценку сходства или различия между вариантами опыта можно проводить по критерию Фишера, критерию Стьюдента или НСР.

Если $F_ф > F_т$, то это позволяет сделать вывод, что внесение больших доз торфа положительно влияет на величину урожая ячменя в агроландшафте и наоборот.

Наиболее распространен в дисперсионном анализе для оценки результатов опыта критерий НСР, алгоритм которого приводим ниже. Вначале определяем среднее квадратическое отклонение из дисперсии, полученной в результате случайного варьирования: $\sigma = \sqrt{\sigma_3^2}$, затем вычисляем обобщенную ошибку среднего: $m_M = \sigma / \sqrt{N_{повт}}$. Поскольку ошибка среднего для всех сравниваемых вариант одна и та же, формула для расчета ошибки разности может быть преобразована: $m_d = \sqrt{2m^2}$.

Таблица 4 **Влияние высоких доз торфа на урожай ячменя**

Вариант опыта	Урожай ячменя по повторностям				Среднее	Прибавка
Контроль (фон)						–
Фон+100 т/га						
Фон+150 т/га						
Фон+300 т/га						
$НСР_{0,95}$, ц/га $НСР_{0,99}$, ц/га						

Практическая работа 3

Тема: **Двухфакторный дисперсионный анализ**

Содержание работы:

По данным влияния метеорологических условий и мелиорации на урожайность биомассы трав в агроландшафте:

1. Построить двухфакторный дисперсионный комплекс.
2. Вычислить девиаты (суммы квадратических отклонений от среднего) и дисперсии результатов опыта по общему варьированию, по факторам и по повторностям и по остаточному варьированию.
3. С помощью критерия Фишера определить влияние исследуемых параметров на биомассу.
4. Оценить результаты эксперимента по критерию наименьшей существенной разности и Стьюдента.
5. Вычислить показатель точности опыта и оценить его достоверность.

Ход работы:

Таблица 1 **Двухфакторный дисперсионный комплекс**

Повторность опыта по фактору II	Биомасса, кг/м^3 Группы по фактору I		$\frac{\sum y_i}{\sum y_i^2}$	$(\sum y_i)^2$	M_y
	сухой	влажный			
Группа по фактору I *(неосушенный агроландшафт)*					
Первая	10	8			
Вторая	12	10			
Третья	10	13			
$\frac{\sum}{\sum}$					
Группа по фактору II *(осушенный агроландшафт)*					
Первая	7	10			
Вторая	8	13			
Третья	8	12			
$\frac{\sum}{\sum}$					
$\frac{\sum x_i}{\sum x_i^2}$				$\sum(\sum y_i)^2 =$	$M_{общ} =$

$(\sum x_i)^2$ M_x			$\Sigma(\sum x_i)^2 =$	

При обработке данных исходной информации (см. табл. 1) порядок расчета не отличается от описанного алгоритма однофакторного дисперсионного комплекса (лаб. работа 3). Дальнейшие расчеты проводятся в следующем порядке.

Общую сумму квадратов отклонений находим по формуле

$$\Theta = \Sigma\Sigma x_i^2, y_i^2 - [(\Sigma\Sigma x_i, y_i)^2 / N],$$

где N – общий объем выборки.

Сумма квадратов отклонений по фактору I вычисляется по формуле

$$\Theta_1 = [\Sigma(\Sigma x_i)^2 - (\Sigma\Sigma x_i, y_i)^2 / n_x] / k_x$$

где n_x – число групп фактора I; k_x – число вариант в каждой отдельной сумме.

Сумма квадратов отклонений по фактору II вычисляется аналогично определению суммы квадратов отклонений по фактору I:

$$\Theta_2 = [\Sigma(\Sigma y_i)^2 - (\Sigma\Sigma x_i, y_i)^2 / n_y] / k_y$$

Сумма квадратов отклонений, вызываемых взаимодействием факторов I и II, определяется следующим образом:

$$\Theta_3 = [\Sigma(\Sigma z_i^2) - (\Sigma\Sigma x_i, y_2)^2 / n_z] / k - \Theta_1 - \Theta_2,$$

где $\Sigma(\Sigma z_i^2)$– сумма квадратов сумм значений вариант по группам выборки комбинационной таблицы; n_z – число сумм вариант по группам; k_z – число слагаемых вариант в каждой группе выборки.

Сумма квадратов отклонений по повторностям Θ_4 определяется по формуле путем подстановки конкретных данных задачи:

$$\Theta_4 = [\Sigma(\Sigma x_i)^2 - (\Sigma\Sigma x_i, y_i)^2 / n_{x,y}] / k_{x,y},$$

где $n_{x,y}$ – число сумм по повторностям; $k_{x,y}$ – число слагаемых в каждой сумме; $\Sigma(\Sigma x_i)^2$ – сумма квадратов сумм исходных данных по повторностям фактора I сверху вниз.

Сумму квадратов отклонений по остаточному варьированию определяем из равенства: $\Theta = \Theta_1 + \Theta_2 + \Theta_3 + \Theta_4 + \Theta_5$

Затем вычисляем число степеней свободы:

-для Θ $\nu = N - 1$;

-для Θ_1 и Θ_2 число степеней свободы равно числу градаций фактора минус единица: $\nu_1 = n_1 - 1$; $\nu_2 = n_2 - 1$;-

-для Θ_3 $\nu_3 = \nu_1 \cdot \nu_2$;

-для Θ_4 число степеней свободы равно числу повторностей минус единица: ν_4; -для Θ_5 этот показатель определяется следующим образом: $\nu_5 = \nu - \nu_1 - \nu_2 - \nu_3 - \nu_4$

Показатели дисперсии (табл. 2) вычисляются путем деления значений сумм квадратов отклонений на соответствующие значения степеней свободы.

Фактический критерий Фишера определяется путем деления каждой из величин дисперсий на значение остаточной.

Если $F_ф > F_т$, то действие данного фактора признается существенным, при $F_ф < F_т$ – несущественным.

Таблица 2 **Результаты двухфакторного дисперсионного анализа**

Варьирование данных	Сумма квадратов отклонений Θ	Степень свободы ν	Дисперсия σ^2	Критерий Фишера	
				$F_ф$	$F_т$
Общее по опыту					4,31
По фактору I					5,99
По фактору II					5,99
По взаимодействию факторов I и II					5,99
По повторностям					5,14
Остаточное					–

Исходя из анализа критерия Фишера необходимо сделать вывод о влиянии исследуемых параметров на биомассу. Оно признается существенным если $F_ф > F_т$. Действие фактора на объект не доказано если $F_ф < F_т$.

Оценку результатов эксперимента можно сделать по критериям НСР и Стьюдента. Для вычисления НСР и t находим ошибку среднего арифметического m_M всего опыта и ошибку разности средних m_d по следующим формулам:

$$m_M = \sqrt{\sigma^2_{ост} / N};\ m_d = \sqrt{2\sigma^2_{ост} / n},$$

где n – численность меньшей из сравниваемых частных групп (в нашем примере обе группы одинаковы и равны шести). Произведите расчет необходимых показателей.

$$НСР = m_d \cdot t_т$$

По критерию Стьюдента сравниваем средние" арифметические данных по осушенному и неосушенному агроландшафту:

$$t = (M_{y,1} - M_{y,2}) / m_{dy}.$$

Сравниваем также средние арифметические по метеорологическим условиям:

$$t = (M_{x,1} - M_{x,2}) / m_{dy}$$

Табличные значения критерия Стьюдента $t_т = 2{,}45$ при $P = 0{,}95$ для $\nu = 6$.

Сделать выводы при использовании критериев Фишера и Стьюдента.

В заключение определить точность опыта, которая равна:

$$p = (m_M / M_{\text{общ}}) \cdot 100\ \%.$$

Практическая работа 4

Тема: Установление тесноты связи и расчет коэффициента корреляции

Содержание работы:

1. Построить график связи между расходами рек Курчум и Ульба (данные государственного водного кадастра)
2. Рассчитать коэффициент корреляции и найти его ошибку с помощью коэффициента Фишера и коэффициента Стьюдента при 5% уровне значимости.
3. Рассчитать уравнение регрессии.

Ход работы:

Таблица 1. Исходные данные для расчета коэффициента корреляции между расходами рек Курчум(y) и Ульба(x)

n	x	y	x·y	x^2	y^2
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					

37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					

Находим среднее арифметическое значение:

$$M_x = \frac{\sum x_n}{n}; \quad M_y = \frac{\sum y_n}{n};$$

Находим среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\overline{x^2} - \overline{x}^2}, \qquad \sigma_y = \sqrt{\overline{y^2} - \overline{y}^2}.$$

Линия регрессии по координатам точек на графике проводится таким образом, чтобы точки в равном количестве находились по обе стороны линии. Более точное значение r получаем расчетным способом, следующим образом как при прямой (r от 0 до 1), так и при обратной (r от 0 до –1) зависимости:

$$r = \frac{\sum (x_i - M_x)(y_i - M_y)}{\sqrt{\sum (x_i - M_x)^2 \sum (y_i - M_y)^2}},$$

где $(x_i - M_x)$, $(y_i - M_y)$ – отклонения значений индивидуальных вариант x_i и y_i от их средних значений M_x и M_y.

Более простой алгебраический расчет коэффициента вариации с учетом объема выборки (n):

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} \quad \text{или} \quad r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}.$$

Исходные данные и суммы по ним получаем из представленной формы:

x_i	x_i^2	$x_i - M_x$	y_i	y_i^2	$y_i - M_y$	xy	$(x_i - M_x)(y_i - M_y)$

Принимается следующая характеристика тесноты корреляционной связи: если r (η) = 0 ± 0,4, то связь считается слабой; от ± 0,4 до ± 0,7 – средняя; от ± 0,7 до ± 1 – сильная; r = ± 1 и η = 1 – связь считается функциональная.

Достоверность вычисленного коэффициента корреляции может быть установлена двумя путями: путем сравнения с табличным значением r; второй путь – через критерий Стьюдента. Если $r_ф > r_{табл}$, то влияние фактора на признак достоверно; если меньше табличного – не достоверно.

При использовании критерия Стьюдента для доказательства достоверности r вначале рассчитывают стандартную ошибку коэффициента корреляции:

$$m_r = \sqrt{(1-r^2)/(N_n-2)},$$

где N_n – число сопряженных пар в сравниваемых выборках.

Значение коэффициента корреляции записывают с учетом его ошибки и уровня значимости: $r_{0,95\ (0,99)} \pm m_r$. Затем вычисляют критерий Стьюдента для коэффициента корреляции: $t_r = r / m_r$

Критерий Стьюдента можно рассчитать иначе: $t_r = r\sqrt{N_n-2}/\sqrt{1-r^2}$

Если вычисленный критерий Стьюдента больше табличного, то зависимость существенна, если меньше – не достоверна. Приближенная оценка статистической достоверности r осуществляется исходя из того, что абсолютное значение r должно превышать ошибку (m_r) в два и более раза.

Линейная регрессия на графике изображается в виде прямой так, чтобы точки эмпирической линии располагались по обе стороны ее и по возможности ближе к ней.

Известно следующее уравнение линейной регрессии: $y = ax + b$

где y – значение зависимой переменной (признак); x – значение независимой переменной (фактор, влияющий на признак); a – коэффициент регрессии, показывающий степень зависимости между переменными (может быть также выражен тангенсом угла наклона линии регрессии к оси абсцисс); b – ордината линии, показывающая смещение начала прямой относительно начала координат.

Для решения поставленной задачи используем *способ координат точек.* Результаты наблюдений наносим на график, затем проводим прямую так, чтобы число точек по обе стороны линии было одинаковым. Для расчета параметров a и b выбираем две точки, которые находятся на прямой или рядом с ней (одну в начале и одну в конце). Подставляя значения переменных в общее уравнение прямой, получаем систему уравнений:

Решаем эту систему относительно a и b. Получив количественное значение параметров a и b, связь между x и y можно выразить конкретным уравнением регрессии.

Уравнение регрессии можно получить также способом наименьших квадратов, используя координаты всех точек. Этот способ заключается в построении такой линии на графике, чтобы сумма квадратов отклонений от нее до точек эмпирической линии регрессии была наименьшей. Для определения параметров a и b составляется система уравнений:

$$\begin{cases} \Sigma y = a\Sigma x + bn; \\ \Sigma xy = a\Sigma x^2 + b\Sigma x. \end{cases}$$

Систему уравнений выводим следующим образом.

$$y_1 = ax_1 + b;$$
$$y_2 = ax_2 + b;$$
$$\dots\dots\dots\dots\dots$$
$$\underline{y_n = ax_n + b}$$
$$\Sigma y = a\Sigma x + bn.$$

Затем каждое исходное уравнение умножаем на соответствующее значение *x;* просуммировав правые и левые части, получим второе уравнение:

$$x_1 y_1 = ax_1^2 + bx_1;$$
$$x_2 y_2 = ax_2^2 + bx_2;$$
$$\dots\dots\dots\dots\dots\dots\dots$$
$$\underline{x_n y_n = ax_n^2 + bx_n}$$
$$\Sigma xy = a\Sigma x^2 + b\Sigma x.$$

Для расчета параметров *a* и *b* составляем табл. 2. Полученные данные подставляем в систему уравнений $\begin{cases} 600 = 830a + 10b; \\ 52102 = 69890a + 830b. \end{cases}$

Решая систему, находим искомые параметры: *a и b*. Подставив полученные показатели в искомое уравнение регрессии, .Хотя значения параметров *a* и *b*, рассчитанные двумя способами, близки между собой, второй способ (наименьших квадратов) более точно определяет положение линии регрессии.

Таблица 2 **Расчет данных для уравнения линейной зависимости**

x	y	xy	x^2	$y'=ax+b$	Расчет критерия χ^2		
					$y-y'$	$(y-y')^2$	$\frac{(y-y')^2}{y'}$
Σ							χ^2 = 30,47

Кроме того, коэффициенты *a* и *b* для уравнения регрессии также могут быть рассчитаны на основе исходных данных (*x, y*) по формулам, которые обеспечивают наименьший квадрат отклонений этих точек от линии регрессии (метод наименьших квадратов):

$$b = \frac{\Sigma x^2 \Sigma y^2 - \Sigma x \Sigma xy}{n_{пар}\Sigma x^2 - (\Sigma x)^2}, \qquad a = \frac{n_{пар}\Sigma xy - \Sigma x \Sigma y}{\Sigma x^2 - (\Sigma x)^2}$$

После составления уравнения регрессии и определения параметров *a* и *b* производим расчет точек *y′* теоретической линии регрессии. Для этого в уравнение регрессии поочередно подставляем значения *x*. Составленные уравнения регрессии можно проверить на точность зависимости между переменными (*x, y)* по коэффициенту точности выравнивания линии r_1, отражающему степень приближения (соответствия) фактических данных наблюдения к вероятным. Этот коэффициент определяем следующим образом:

$$r_1 = \sqrt{\frac{\Sigma\alpha^2 - \Sigma\beta^2}{\Sigma\alpha^2}} = \sqrt{\frac{\Sigma(y_ф - M_ф)^2 - \Sigma(y_ф - y_в)^2}{\Sigma(y_ф - M_ф)^2}},$$

где ($y_ф$ – $M_ф$) = α – отклонение индивидуальных вариант от общего среднего арифметического по *y;* ($y_ф$ – $y_в$) = β – отклонение индивидуальных экспериментальных вариант по *y* от расчетных по уравнению.

Принято считать: если $r_1 > 0{,}95$, то уравнение регрессии соответствует более точному положению линии на графике. При $r_1 < 0{,}95$ необходимо найти другую математическую зависимость. В приведенном примере $r_1 = 0{,}88 < 0{,}95$, поэтому следует подобрать другую математическую зависимость. Такие же выводы получены при проверке на точность зависимости между переменными по критерию хи-квадрат. Оба критерия оценки (χ^2, r_1) на точность выравнивания линии уравнения регрессии используются и для других форм регрессионной зависимости.

Ошибку уравнения регрессии можно определить по формуле

$$m = \sqrt{\frac{\Sigma(y - y')^2}{n - k}} = \sqrt{\frac{\sum(y_ф - y_в)^2}{n - k}},$$

где *n* – число точек линии регрессии; *k* — число коэффициентов в уравнении регрессии (два плюс свободный член уравнения).

Практическая работа 5

Тема: Вычисление коэффициента ранговой корреляции Спирмена

Содержание работы:

1. Рассчитать коэффициент ранговой корреляции К.Спирмена (r_c).

Коэффициент ранговой корреляции К.Спирмена (r_c) относится к непараметрическим критериям связи. Коэффициент Спирмена представляет собой следующее соотношение:

$$r_c = 1 - \frac{6\Sigma(x'-y')^2}{N_n^{\ 3} - N_n}, \text{ или } r_c = 1 - \frac{6\Sigma(d^2)}{N_n^{\ 3} - N_n},$$

где d – разность между сопряженными рангами; x' – величины рангов, заменяющие фактические варианты или качественные признаки по аргументу x; y' – величины рангов, заменяющие фактические варианты или качественные признаки по функции y; $N_п$ – количество сопряженных пар.

Ранги присваиваются по возрастающей величине. Результаты используются для расчета рангового коэффициента корреляции по формуле.

Для получения необходимых показателей при расчете рангового коэффициента корреляции составляем таблицу 1. Вычисляем разность между парными рангами (x'–y'), которые возводим в квадрат и суммируем.

Расчет рангового коэффициента корреляции

x'	y'	x'–y'	$(x'-y')^2$

Коэффициент ранговой корреляции имеет значения между +1 и –1, что больше соответствует фактической связи между признаками.

2. *Установить достоверность* коэффициента ранговой корреляции К.Спирмена (r_c).

Достоверность полученного рангового коэффициента можно установить аналогично достоверности коэффициента корреляции.

Если ранговый коэффициент корреляции $r_ф > r_т$ при P= 0,90 для ν = n , можно сделать вывод, что влияние исследуемых факторов достоверно и положительно.

Данные для вычисления коэффициента корреляции
(Выбираете один вариант)

Вариант 1.

n	x	y	n	x	y
1	400	8	5	800	58
2	500	6	6	900	80
3	600	16	7	1000	90
4	700	34	8	1200	95

Вариант 2

n	x	y	n	x	y
1	26	37	7	52	56
2	30	46	8	58	62
3	33	48	9	65	65
4	36	49	10	68	67
5	54	50	11	70	74
6	46	54	12	74	79

Вариант 3

n	x	y	n	x	y
1	30	23,5	6	40	28,8
2	35	23,7	7	45	33,5
3	35	24,0	8	37	27,7
4	38	26,7	9	35	23,0
5	29	24,3	10	40	29,4

Вариант 4

n	x	y	n	x	y
1	74	18,9	6	70	14,5
2	78	27,3	7	78	18,0
3	72	13,0	8	62	16,1
4	64	18,1	9	63	16,0
5	64	15,7	10	62	13,3

Вариант 5

n	x	y	n	x	y
1	64	13,4	6	33	10,0
2	58	11,8	7	78	21,0
3	64	15,7	8	70	17,0
4	54	15,8	9	45	14,0
5	20	8,0	10	33	14,0

Практическая работа 6

Тема: Алгоритм проведения корреляционного и регрессионного анализов в Microsoft Office Excel 2003

Методика выполнения:

Проверим зависимость между баллом пашни (*x*) и урожайностью многолетних трав (*y*), для чего набираем в ячейках A2:K3 следующие данные:

x	43	42	38	36	33	45	40	45	36	32
y	33,2	18,6	28,4	26,5	30,9	31,8	32,4	30,6	26,8	24,4

Строим точечную диаграмму: выделяем набранную таблицу (ячейки A2–K3) и жмем на пиктограмму  на панели инструментов или *Вставка–Диаграмма,* в закладке *Стандартные* выбираем *Точечная* и первый сверху из имеющихся примеров жмем *Далее,* в закладке *Диапазон данных* отмечаем *Ряды в строках – Далее.* В закладке *Заголовки* в окошке *Ось X (категорий)* набираем «Балл пашни» (может отличаться для различных индивидуальных заданий, в этом случае пишется название первого сравниваемого параметра), в окошке *Ось Y (значений)* «Урожайность многолетних трав», в закладке *Легенда* снимаем галочку с показателя «*Добавить легенду»– Далее – Поместить диаграмму на имеющемся листе – Готово*.

Добавляем линию тренда, для чего кликаем на маркере точки данных правой клавишей и выбираем пункт *Добавить линию тренда* (см. рис. 1).

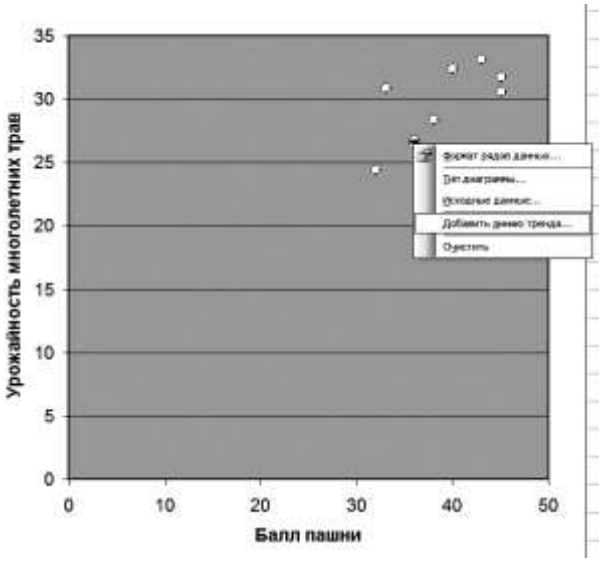
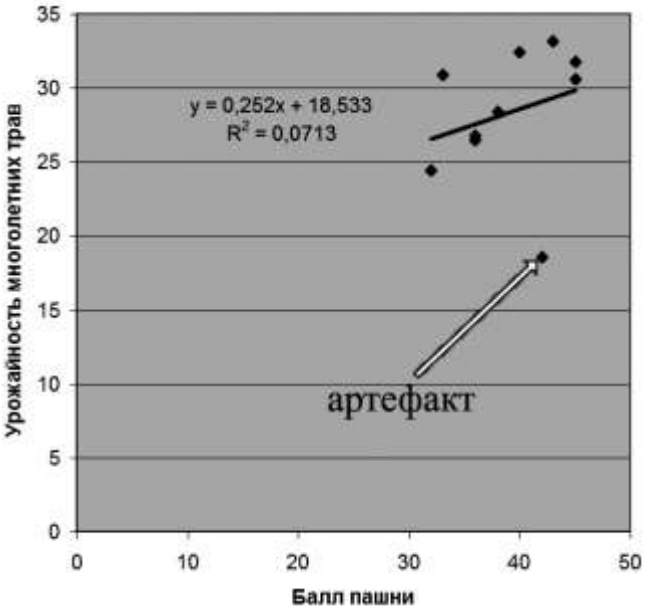


Рис. 1

В закладке *Тип* выбирается *Линейная,* в закладке *Параметры* отмечаются пункты *показывать уравнение на диаграмме* и *поместить на диаграмму величину достоверной аппроксимации – ОК*. В итоге будет построена линия тренда и составлено уравнение линейной регрессии. Находим артефакты – значения, которые сильно отдалены от линии тренда и не вписываются в общую картину (рис. 2). Более правильно выявлять артефакт через расчеты приведенные в п. 1.2. Удаляем эти значения из таблицы данных (в указанном примере случае очищаются от цифр ячейки C2:C3), MS Excel автоматически пересчитает уравнение регрессии. В некоторых случаях (при нелинейной корреляции), можно использовать другие варианты линий тренда, например логарифмическую, степенную или экспоненциальную.



Рассчитываем коэффициент корреляции установив курсор в клетку B5 используя команду КОРРЕЛ: *Вставка – Функция* (или иконка $\mathbf{f_x}$ на панели инструментов) – выбираем в категориях *Статистические* – функцию КОРРЕЛ – сворачиваем появившееся окно нажатием на кнопку напротив поля *Массив 1*. Выделяем ячейки со значениями ***x*** (B2:K2), далее в поле *Массив 2* выделяем ячейки со значениями ***y*** (B3:K3), разворачиваем окно, нажав на эту же кнопку и жмем *ОК*.

Оцениваем значимость коэффициента корреляции (***r***) по критерию Стьюдента по формуле $t_r = \sqrt{N-2} / \sqrt{1-r^2}$ и сравниваем с табличным (критическим) значением, если фактическое значение больше критического, то корреляционная связь существенна, если меньше – недостоверна (вид формул на рис. 3).

	A	B
1	Между балом пашни и урожайностью	многолетних трав
2	x	43
3	y	33,2
4		
5	Коэф. корреляции	=КОРРЕЛ(B2:K2;B3:K3)
6	Критерий Стьюдента	=B5*КОРЕНЬ(СЧЁТ(B2:K2)-2)/КОРЕНЬ(1-B5*B5)
7	Критическое значение критерия Стьюдента	=СТЬЮДРАСПОБР(0,05;СЧЁТ(B2:K2)-2)

Рис.3

Регрессионный анализ проводится с помощью надстройки «*Пакет анализа*», для последовательность команд *Сервис – Анализ данных – Регрессия,* в поле «*Входной интервал*» указываем значения для Y и X (A3:K3 и A2:K2 соответственно), в «*Параметрах вывода*» выбираем «*Выходной интервал*» и указываем там ячейку на этом же листе, отмечаем параметры «*Уровень надежности*» (значение можно изменять, в нашем случае указываем 95%), нажимаем [OK]. Если удалялся артефакт, то необходимо скопировать первоначальные значения в другие ячейки, поскольку значения во входном интервале должны быть непрерывными.

Задание.

Проверить зависимость между баллом бонитета (*x*) и урожайностью озимой пшеницы по хозяйствам Талды-Курганской области (*y*)

Данные:

n	Балл бонитета (x)	Урожайность (y)
1	74	18,9
2	78	27,3
3	72	13,0
4	64	18,1
5	64	15,7
6	70	14,5
7	78	18,0
8	62	16,1
9	63	16,0
10	62	13,3
11	64	13,4
12	58	11,8
13	64	15,7
14	54	15,8
15	20	8,0
16	33	10,0
17	78	21,0
18	70	17,0
19	45	14,0

20	33	14,0
21	21	9,0

Практическая работа 7

Тема: Вычисление коэффициента линейной корреляции взвешанных рядов

Содержание работы:

1. Составить взвешанные вариационные ряды стоков р.Черный Иртыш у с. Буран и р.Ульба у с. Ульба Перевалочное.
2. Построить корреляционную таблицу частот, сопоставить величины стока.
3. Методом условных средних вычислить коэффициент корреляции между годовыми величинами стока рек Иртыш и Ульба.
4. С помощью критерия Стьюдента определить достоверность оценки коэффициента корреляции при уровне значимости 5%.
5. Оценить достоверность с помощью критерия Фишера.
6. Вычислить ср.годовые величины стока рек и коэффициенты корреляции. Оценить точность вычисленных величин.

Ход работы:

1. Составить таблицу взвешанные вариационные ряды стоков р.Черный Итрыш у с. Буран и р.Ульба у с. Ульба Перевалочное

Таблица 1. Годовые расходы р.Черный Иртыш у с. Буран (х) и р.Ульба у с. Ульба Перевалочное (у)

n	x	y	x↑	y↑
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				

Таблица 2. Корреляционная таблица годовых расходов воды р.Черный Иртыш у с. Буран (х) и р.Ульба у с. Ульба Перевалочное (у)

y / x	i =							f_x	A_x	$f_x \cdot A_x$	$f_x \cdot A_x^2$
i =	y / x										
f_y								$\sum$		$\sum$	$\sum$

A_y								
$f_y \cdot A_y$								$\sum$
$f_y \cdot {A_y}^2$								$\sum$
$f_{yx} A_y A_x$								$\sum$

$$A_x = \frac{(x_i - A_x)}{x_i}; \quad A_y = \frac{(y_i - A_y)}{y_i};$$

$$k = 1 + 3.3 \lg n \; ; \qquad i = \frac{x_{max} - x_{min}}{k},$$

где k - количество классов в ряду, i – *величина классового интервала,* x_{max}, x_{min} - *крайние значения признаков,* n – *объем выборки.*

$$b_x = \frac{\sum f_x \cdot A_x}{n}; \qquad b_y = \frac{\sum f_y \cdot A_y}{n};$$

Вычислить средние величины:

$$\bar{x} = A_x + x_i \cdot b_x; \qquad \bar{y} = A_y + y_i \cdot b_y$$

2 Вычислить среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\frac{\sum f_x \cdot {A_x}^2}{n} - {b_x}^2}; \qquad \sigma_y = \sqrt{\frac{\sum f_y \cdot {A_y}^2}{n} - {b_y}^2};$$

3. Вычислить коэффициент корреляции:

$$r = \frac{\sum f_{xy} \cdot A_x \cdot A_y - n \cdot b_x \cdot b_y}{n \cdot \sigma_x \cdot \sigma_y};$$

4. Ошибка коэффициента корреляции: $m_r = \sqrt{(1 - r^2)/(n-1)}$; С помощью критерия Стьюдента определить достоверность оценки коэффициента корреляции при уровне значимости 5%: $t_r = r / m_r$, $t_m = 2{,}01$

Если вычисленный критерий Стьюдента больше табличного, то зависимость существенна, если меньше – не достоверна.

5. Расчет коэффициента корреляции по преобразованной величине Z

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r};$$

Оценка достоверности с помощью критерия Фишера: $F_z = Z\sqrt{n-3}$; $F_{zm} = 2{,}80$

Практическая работа 8

Тема: Расчет экономических характеристик производственной функции

Содержание работы.

1. Рассчитать некоторые экономические характеристики производственной функции:

$$y = 3{,}19 + 0{,}126x_1 + 0{,}81x_2 + 0{,}102x_3,$$

где y – плотность поголовья коров на 100 га сельхозугодий; x_1 – площадь кормовых угодий (в % от общей площади сельхозугодий); x_2 – стоимость животноводческих построек (тыс.тенге на 100 га сельхозугодий); x_3 – площадь смытых земель (в % от общей площади сельхозугодий).

а) Дополнительный продукт по отдельным факторам: $D_i = \frac{\partial y}{\partial x_i}$;

- по площади кормовых угодий (x_1);
- стоимость животноводческих построек (x_2);
- площадь смытых земель (x_3).

Производная D_i характеризует скорость (темп) изменения показателя эффективности «в данной точке» при изменении i-го фактора и заданных значениях других производственных факторов.

б) Коэффициент эластичности:

- по площади кормовых угодий (x_1);
- стоимость животноводческих построек (x_2);
- площадь смытых земель (x_3).

$$E_i = \left(\frac{\partial y}{y}\right) \Big/ \left(\frac{\partial x_i}{x_i}\right) = \frac{D_i}{\Pi_i};$$