

Міністерство освіти і науки України
Державний ВНЗ
НАЦІОНАЛЬНИЙ ГІРНИЧИЙ УНІВЕРСИТЕТ



ФАКУЛЬТЕТ МЕНЕДЖМЕНТУ

Кафедра прикладної економіки та підприємництва

Завдання для проведення практичних занять
з початкової дисципліни
«Проектування підприємницької діяльності»

Галузь знань 07 Управління та адміністрування

.....

Спеціальність 076 Підприємництво, торгівля та
..... біржова діяльність

Дніпро, 2017

Завдання для проведення практичних занять з *початкової дисципліни* «Проектування підприємницької діяльності» для студентів спеціальності 076 Підприємництво, торгівля, біржева справа / Упоряд.: В.І. Прокопенко. – Д.: Національний гірничий університет, 2017. – 21 с.

Упорядники:

В.І. Прокопенко, д-р техн. наук, проф.

Затверджено методичною комісією в галузі знань 07 Управління та адміністрування (протокол № 1 від 6.04.17) за поданням кафедри прикладної економіки та підприємництва (протокол № 13 від 21.03.17).

Відповідальний за випуск – професор кафедри прикладної економіки та підприємництва д-р техн. наук, проф. В.І. Прокопенко.

ЗАВДАННЯ № 1

*Визначення необхідного числа спостережень***Постановка завдання**

1. Запропонуйте й опишіть приклад досліджень шляхом проведення експериментів, спостережень або вимірювань з метою встановлення середньоарифметичного значення будь-якого техніко-економічного показника діяльності організації (підприємства, виробничого підрозділу, механізму тощо).
2. Викладіть методику встановлення необхідного обсягу статистичної вибірки (кількості спостережень).
3. Обґрунтуйте вихідні дані, потрібні для встановлення обсягу вибірки.
4. Встановіть необхідну кількість спостережень техніко-економічного показника, що досліджується.
5. Сформулюйте висновок стосовно отриманого обсягу статистичної вибірки.

Методика вирішення

Встановлюється мета дослідження та здійснюється вибір об'єкта, предмета та ситуації дослідження. Обирається певний техніко-економічний показник діяльності об'єкта дослідження, щодо якого існує вірогідність суттєвого коливання його величини відносно математичного сподівання при проведенні за рівних умов вимірів, спостережень або експериментів. Після цього розглядається наявна сукупність однорідних даних досліджуваної ознаки, з якої, за визначеним принципом (наприклад – випадковим чином, систематично, квотно тощо) формується вибірка сукупність (вибірка). Вибірка має адекватно відображати особливості розподілу даних генеральної сукупності. З цією метою вона перевіряється на достатність даних. Необхідно встановити мінімальне число показників, за якого вибірка не втрачає репрезентативності. Після цього виконується статистична обробка даних, згідно з метою дослідження.

Методика складається з наступних етапів:

1. Розраховується ймовірність появи випадкового відхилення за відношенням:

$$\varepsilon = \frac{\Delta x}{S},$$

де Δx – припустима величина випадкового відхилення вибірки;

S – середньоквадратичне відхилення вибірки.

2. Приймається коефіцієнт надійності експерименту (α), виходячи з умов задачі та призначення експериментальних даних. Як-правило, для задач економічного характеру α приймається рівним 0,9.

3. За табл.1 (дод.1) знаходиться, згідно зі значеннями ε та α , число мінімально необхідних спостережень досліджуваної ознаки.

4. Формулюється висновок стосовно отриманого обсягу статистичної вибірки.

Приклад виконання завдання

1. Досліджується перемінна частка витрат на маркетингові дослідження у кошторисі витрат на стратегічно важливі інвестиційні проекти. Для цього встановлюється необхідний обсяг статистичної вибірки.

2. Викладається методика.

Визначають основні положення для встановлення кількості даних, які необхідно зібрати у результаті спостережень за витратами на маркетингові дослідження. Положення наведені вище у пп. 1- 4.

3. Приймаємо середньоквадратичне відхилення даних вибірки, таким, що дорівнює 5,5%. Випадкове відхилення представлених значень, згідно з метою дослідження, дорівнює 1,5%. Коефіцієнт надійності експерименту (α) прийнято рівним 0,9. Коефіцієнт вірогідності появи випадкового відхилення $E = 1,5/5,5 = 0,27$

4. Визначаємо необхідну кількість спостережень за витратами на маркетингові дослідження згідно з наступними даними: $E = 0,27$, $\alpha = 0,9$. За табл.1 (дод.1) число необхідних спостережень дорівнює 32.

5. Висновок: Таким чином, вибіркова сукупність повинна мати не менше 32 спостережень. Цей обсяг даних дозволить розрахувати таку середньоквадратичну величину, відносно якої в інтервалі припустимого випадкового відхилення $\pm 1,5\%$ будь-який результат спостереження буде потрапляти з вірогідністю $\alpha = 0,9$.

Таблиця 1

Необхідне число вимірів для отримання випадкової похибки ε з надійністю α

$\varepsilon = \frac{\Delta x}{S}$	α					
	0,5	0,7	0,9	0,95	0,99	0,999
1,0	2	3	5	7	11	17
0,5	3	6	13	18	31	50
0,4	4	8	19	27	46	74
0,3	6	13	32	46	78	127
0,2	13	29	70	99	171	277
0,1	47	169	273	387	668	1089
0,05	183	431	1084	1540	2659	4338
0,02	4543	10732	27161	38416	66358	108307

ПОСТАНОВКА ЗАВДАННЯ № 2

Перевірка на нормальність розподілу вибіркової сукупності

1. Запропонуйте й опишіть приклад досліджень шляхом проведення експериментів, спостережень або вимірювань.

2. Складіть статистичну вибірку для перевірки на нормальність розподілу будь-якого техніко-економічного показника діяльності організації (підприємства, виробничого підрозділу, механізму тощо). Обсяг статистичної ви-

бірки – не менш 40 спостережень (вимірювань).

3. Викладіть методику перевірки статистичної вибірки на розподіл за нормальним законом.

4. Здійсніть перевірку складеної вибірки за поз. 2 на нормальність розподілу.

5. Сформулюйте висновок стосовно одержаного результату перевірки.

Методика вирішення

1. Розглядають випадковий техніко-економічний показник, закон розподілу якого є невідомим, але припускається, що він розподілений за нормальним законом. З метою підтвердження цього вибірка перевіряється на нормальність розподілу. Тоді з'являється можливість управління даною сукупністю показників.

2. За визначеним принципом утворюють вибірку сукупність достатнього обсягу. Значення сукупності упорядковують, виконавши команду *Дані / Сортування (сортувати за зростанням)*

Методика складається з наступних етапів:

а) вибірку розбивають на рівні інтервали, величина яких визначається за виразом:

$$h = \frac{X_{\max} - X_{\min}}{1 + 3,2Lgn}, \quad (2.1)$$

де X_{\max} – максимальне значення вибірки;

X_{\min} – мінімальне значення вибірки;

n – число спостережень.

Нижньою границею початкового інтервалу буде мінімальне значення вибірки, верхньою – мінімальне, збільшене на величину (крок) інтервалу. Останнє, в свою чергу, буде нижньою границею наступного інтервалу, а верхня – визначатиметься кроком інтервалу. Останній інтервал має вмещувати максимальне значення вибіркової сукупності даних.

б) за кожним інтервалом знаходять його середнє значення x_i як суму верхньої та нижньої границь відповідного інтервалу, поділену навпіл. Виділивши поле, що дорівнює числу інтервалів, та скориставшись командою *Вставка / Функція / Статистические / Частота* вбудованих функцій Excel знаходять частоту появи значень кожного інтервалу m_{ei} , зазначивши у діалоговому вікні „*Аргументы функции*” у полі „*Массив данных*” массив даних вибірки, а у полі „*Массив интервалов*” – верхні границі інтервалів сукупності даних. Для виконання команди одночасно натискають клавіші <Ctrl> + <Shift> + <Enter>.

в) розраховують середнє значення всієї сукупності даних x_i та її середньоквадратичне відхилення δ за допомогою команд, відповідно, *Вставка / Функція / Статистические / СРЗНАЧ* та *Вставка / Функція / Статистические / СТАНДОТКЛП*, де у полі діалогових вікон програми зазначають весь діапазон даних вибірки;

з) для кожного інтервалу значень визначають теоретичну (вирівнюючу) частоту за виразом:

$$m_{T_i} = \frac{nh}{\delta_e} \cdot \varphi(t_i), \quad (2.2)$$

де $\varphi(t_i)$ – табличне значення функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ вірогідності появи теоретичного значення вибірки (середнього значення i -го інтервалу).

Параметр t_i як кількісне вираження вірогідності появи середнього значення x_i виділеного i -го інтервалу визначають за формулою:

$$t_i = \frac{x_i - \bar{x}}{\delta}. \quad (2.3)$$

д) встановлюють відносні емпіричні частоти за виразом:

$$m_{ei}' = \frac{m_{ei}}{n_e}, \quad (2.4)$$

де m_{ei}' – емпірична частота i -го інтервалу;

n_e – число спостережень;

е) визначимо відносні теоретичні частоти за виразом:

$$m_{Ti}' = \frac{m_{Ti}}{\sum n_T}, \quad (2.5)$$

де m_{Ti} – теоретична частота i -го інтервалу;

$\sum n_T$ – сума теоретичних частот;

ж) розраховують накопичені емпіричні $F(m_{ei}')$ та теоретичні $F(m_{Ti}')$ частоти як суму відповідної відносної частоти i -го інтервалу та відносних частот попередніх інтервалів;

з) визначають різницю між накопиченими емпіричними та теоретичними частотами за кожним інтервалом та встановлюють серед них максимальне відхилення D_{\max} ;

і) встановлюють фактичний критерій відхилення емпіричного розподілу від теоретичного за наступної формулою:

$$\lambda_\phi = D_{\max} \cdot \sqrt{n}. \quad (2.6)$$

к) визначають граничне значення відхилення λ_k емпіричної функції від теоретичної за накопиченими частотами згідно з критерієм Колмогорова за табл. 2.1, виходячи з рівня значущості результатів розрахунків k .

Таблиця 2.1

Граничний критерій відхилення емпіричного розподілу від теоретичного

k	10	5	2	1	0,5	0,1
λ_k	1,224	1,358	1,517	1,627	1,731	1,950

л) порівнюють значення λ_{ϕ} з λ_{κ} . Якщо $\lambda_{\phi} \leq \lambda_{\kappa}$, то роблять висновок, що емпіричний розподіл даних не суперечить нормальному розподілу.

Приклад виконання завдання

1. Досліджується перемінна частка витрат на маркетингові дослідження у кошторисі витрат на стратегічно важливі інвестиційні проекти (табл. 2.2), що має достатній обсяг спостережень, але розподіл випадкової величини є невідомим. Необхідно перевірити вибірку на нормальність розподілу.

2. Впорядкуємо розміщення даних у зростаючому порядку за допомогою програми *Майстер функцій* (п. 2).

3. Викладається методика перевірки вибірки на нормальність розподілу.

4. Здійснюється перевірка вибірки на нормальність розподілу:

а) розіб'ємо вибірку на інтервали, визначивши їх крок за формулою (2.1): $h = 2,4$. Результат розрахунків представимо у табл. 2.3, гр. 1 та на рис. 2.1.

Таблиця 2.2

Вихідні дані перевірки на нормальність розподілу

№ проекту	Частка маркетингових досліджень, %	№ проекту	Частка маркетингових досліджень, %	№ проекту	Частка маркетингових досліджень, %
1	15,5	12	16,8	23	21,9
2	16,9	13	17,2	24	18,5
3	16	14	18	25	14,4
4	16,2	15	16,7	26	20
5	15,2	16	16,8	27	16,7
6	16	17	17,9	28	21,3
7	10,2	18	23	29	21,8
8	18,2	19	23,6	30	16,8
9	16,3	20	24,3	31	22
10	17,9	21	19,2	32	21,6
11	21,1	22	15,3		

б) знаходимо середню частку маркетингових досліджень x_i та частоту появи значень кожного інтервалу m_{ei} згідно з викладеною вище методикою. Результати розрахунків представимо у табл. 2.3, гр. 2 та 3.

Таблиця 2.3

Розрахунок накопичених частостей

	Середнє	Частота	Відносна частість	Відносна накопичена частість	Різниця

Інтервал	значення, x_i							
		m_{e_i}	m_{T_i}	m_{e_i}'	m_{T_i}'	$F(m_{e_i}')$	$F(m_{T_i}')$	(гр.7-гр.8)
1	2	3	4	5	6	7	8	9
10,2-12,6	11,4	1	0,84	0,03	0,04	0,03	0,04	- 0,01
12,7-15,0	13,8	1	3,64	0,03	0,15	0,06	0,19	- 0,13
15,1-17,4	16,2	14	8,26	0,44	0,33	0,5	0,52	- 0,02
17,5-19,8	18,6	6	3,29	0,19	0,13	0,69	0,65	0,04
19,9-22,2	21	7	6,51	0,22	0,27	0,91	0,92	- 0,01
22,3-24,6	23,4	3	2,23	0,09	0,08	1,00	1,00	0,00
Σ		32	24,79	1	1			

в) знаходимо середнє значення вибіркової сукупності \bar{x} та середньоквадратичне відхилення вибірки δ за допомогою програми *Майстер функцій*:

$$\bar{x}=18,2, \delta= 3,05;$$

г) для кожного інтервалу часток маркетингових досліджень встановлюємо теоретичні частоти за формулою (2.2), (табл. 2.3, гр. 4) з попередньо визначеним параметром t_i згідно з формулою (2.3);

д) визначимо відносні емпіричні частоти m_{e_i}' за формулою (2.4) (табл. 2.3, гр. 5):

е) визначимо відносні теоретичні частоти m_{T_i}' за формулою 2.5 (табл. 2.3, гр. 6);

ж) розрахуємо, відповідно, накопичені відносні емпіричні $F(m_{e_i}')$ та теоретичні частоти $F(m_{T_i}')$ (табл. 2.3, гр. 7 - 8) згідно з пунктом ж методики;

з) визначимо різницю накопичених емпіричних та теоретичних частот часток маркетингових досліджень за кожним інтервалом та встановимо їх максимальне відхилення D_{\max} (табл. 2.3, гр. 9): $D_{\max} = 0,13$;

і) встановимо фактичний критерій відхилення емпіричного розподілу від теоретичного за Колмогоровим згідно з формулою (2.6): $\lambda_{\phi} = 0,735$;

к) за табл. 2.1 визначаємо граничний критерій Колмогорова відповідності емпіричних даних нормальному розподілу. За рівнем значущості 5%, що відповідає умові задачі, $\lambda_{\kappa} = 1,358$;

л) порівнюємо розрахункове та табличне значення критерію Колмогорова

ва: $\lambda_{\phi} = 0,735 < \lambda_{\kappa} = 1,358$.

5. Висновок: оскільки розрахункове значення критерію відповідності емпіричного розподілу 0,735 менше теоретичного критерію Колмогорова 1,358, то можна стверджувати, що сукупність часток маркетингових даних, яка досліджується, підкоряється нормальному закону розподілу, а результати обстежень (зазначені частки) є випадковими величинами й управління ними потребує спеціальних засобів.

Перевірка на нормальність розподілу окремих результатів спостережень

1. Запропонуйте й опишіть приклад досліджень шляхом проведення експериментів, спостережень або вимірювань.

2. Складіть статистичну вибірку для перевірки на нормальність розподілу будь-якого техніко-економічного показника діяльності організації (підприємства, виробничого підрозділу, механізму тощо). Обсяг статистичної вибірки – не менш 40 спостережень (вимірювань).

3. Викладіть методику перевірки статистичної вибірки на розподіл за нормальним законом.

4. Здійсніть перевірку складеної вибірки за поз. 2 на нормальність розподілу.

5. Сформулюйте висновок стосовно одержаного результату перевірки.

Методика вирішення

Встановлюється техніко-економічний або інший показник, що викликає дослідницький інтерес. За визначеним принципом, шляхом проведення експериментів, спостережень або вимірювань, фіксується його абсолютне значення, що є змінним, складається сукупність цих значень, яка підлягає певному закону розподілу. Перевіряється, чи є отримана сукупність даних техніко-економічного аналізу розподіленою за нормальним законом. Цьому закону підкоряється поява незалежних випадкових величин, якими можна керувати на підставі закономірностей їх виникнення.

Із можливих значень досліджуваного показника випадковим чином формують вибірку сукупність достатнього обсягу, яка віддзеркалюватиме закономірність появи змінної величини згідно з вимогами отримання заданої точності розрахунків.

Методика складається з наступних етапів:

1. Визначають значення вибірки, що найбільш відрізняється від решти показників (x_i).

2. Розраховують математичне сподівання вибірових значень, як середньоарифметичне всіх даних вибірки (\bar{x}).

3. Встановлюють середньоквадратичне відхилення вибірки від математичного сподівання за виразом:

$$\delta = \sqrt{\left(\frac{\sum x_i^2}{n} - \bar{x}^2 \right)}, \quad (2.1)$$

де $\sum x_i^2$ – сума квадратів значень вибірки;

\bar{x} – математичне сподівання вибірки;

n – обсяг вибірки.

4. Підставляють отриманні значення у пп. 1-3 у наступний вираз:

$|x_i - \bar{x}| < 3\delta$, що відображає ймовірність розподілу випадкової величини за нормальним законом згідно з правилом „трьох сігм”. Якщо нерівність виконується, то вибірка розподілена за нормальним законом.

5. Формулюють висновок стосовно перевірки даних на нормальність розподілу.

Приклад виконання завдання

1. Досліджується перемінна частка витрат на маркетингові дослідження у кошторисі витрат на стратегічно важливі інвестиційні проекти (табл. 2.1), що має достатній обсяг спостережень, але розподіл випадкової величини є невідомим. Необхідно перевірити вибірку на нормальність розподілу.

Таблиця 2.1

Частка витрат на маркетингові дослідження

№ проекту	Частка витрат, %	№ проекту	Частка витрат, %	№ проекту	Частка витрат, %
1	15,5	12	16,8	23	21,9
2	16,9	13	17,2	24	18,5
3	16	14	18	25	14,4
4	16,2	15	16,7	26	20
5	15,2	16	16,8	27	16,7
6	16	17	17,9	28	21,3
7	10,2	18	23	29	21,8
8	18,2	19	23,6	30	16,8
9	16,3	20	24,3	31	22
10	17,9	21	19,2	32	21,6

11	21,1	22	15,3		
----	------	----	------	--	--

2. Викладається методика виконання завдання, основні положення якої наведені вище у пп. 1- 4.

3. Знаходимо значення вибірки, що може не відповідати нормальному розподілу. Для отриманої сукупності часток витрат на маркетингові дослідження $x_i = 10,2$. Визначимо математичне сподівання вибірових значень \bar{x} . Воно дорівнює 18,2. Середньоквадратичне відхилення вибірки (δ) встановлюємо за виразом (2.1). З розрахунків витікає, що $\delta = 5,5$. Перевіряємо відповідність нормальності розподілу значень, наведених у табл. 2.1, підставивши результати розрахунків в умову нормального розподілу випадкової величини:

$$| 10,2 - 18,2 | < 3 \cdot 5,5 ;$$

$$8 < 16,5.$$

4. Так як нерівність виконується, то можна стверджувати наступне. Вірогідність появи абсолютного значення досліджуваної величини 10,2% не перевищує граничну вірогідність появи випадкової величини нормального розподілу даних, а відтак, вибіркова сукупність часток витрат на маркетингові дослідження розподілена за нормальним законом.

ЗАВДАННЯ № 3

Планування експериментів

Постановка завдання

1. Запропонуйте й опишіть приклад досліджень шляхом проведення експериментів, спостережень або вимірювань.

2. Розробіть рандомізований план проведення досліджень для встановлення кореляційної залежності будь-якого техніко-економічного показника діяльності організації (підприємства, виробничого підрозділу, механізму тощо) від однієї з керованих факторних перемінних, які впливають на цей показник. Виключіть при цьому вплив трьох зовнішніх незалежних перемінних. План має бути складений у вигляді двох квадратів 4x4.

3. Опишіть, яким чином слід проводити спостереження за Вашим планом експерименту.

4. Напишіть результати спостережень та розрахуйте середньо статистичні дані для визначення зазначеної вище залежності.

Методика вирішення

Необхідно провести дослідження з метою встановлення кількісної залежності між будь-якими параметрами процесу, що вивчається. При цьому визначають техніко-економічний показник, який характеризує ефективність діяльності об'єкта. Його розрахунок має здійснюватися за умови створення найбільш типових умов функціонування об'єкта. На зазначений показник впливають як фактори, якими можна керувати (внутрішні залежні змінні),

так і фактори, що не є керованими (зовнішні незалежні змінні). Якщо на досліджуваний показник впливає одна зовнішня змінна, то її вплив усувається шляхом усереднення отриманих у ході дослідження спостережень. При наявності більшої кількості незалежних змінних, які впливають на визначений параметр, складають рандомізований план проведення спостережень шляхом випадкового поєднання незалежних змінних з керованим фактором. Чим менше послідовностей (закономірностей) розподілу вихідних даних має такий план, тим досконалішим він вважається.

Методика виконання завдання складається з наступних етапів:

1. Встановлюють залежність, яку треба встановити, та визначають склад найбільш впливових незалежних випадкових факторів, що можуть її порушувати. Вирішують план проведення спостережень таким чином, щоб ці спостереження були проведені в умовах різних зовнішніх змінних. Математичне сподівання досліджуваного показника визначають як середньо арифметичне отриманих результатів експерименту.

2. Для встановлення статистичної залежності $y_i = f(x_i)$ достатньо мати 6...8 значень показника, що вивчається, та відповідні їм значення керованого фактору. Тому достатньо побудувати рандомізований план у вигляді двох квадратів розмірністю 4 x 4, в якому розміщують вісім величин залежного фактору, що у табличному вигляді для збалансування умов експерименту по чергово поєднуються з іншими незалежними змінними.

3. Описують, яким чином слід проводити спостереження за Вашим планом з визначенням переліку незалежних змінних та керованої змінної факторних ознак.

4. На підставі складеного рандомізованого плану дослідження встановлюють середні значення результатів експерименту за ознакою, що вивчається.

Приклад виконання завдання

1. Потрібно встановити залежність $y_i = f(x_i)$, де y_i - обсягу перевезень гірських порід, а x_i - відстань транспортування, що дозволить визначити норму виробітку водія автосамоскида для різної відстані перевезень. Вимірюється обсяг перевезень y_i при різних відстанях (внутрішня регульована змінна $x_i = 1,2; 1,8; 2,5; 2,8; 3,0; 3,5; 4,2$ і $5,5$ км), де як незалежні зовнішні змінні прийняті день тижня (ПН,ВТ,СР,ЧТ), водій (А,В,С,Д) та автомобіль (Z,Y,X,W).

2. Кожна регульована змінна поєднується у рандомізованому плані з іншими незалежними факторами так, щоб в сукупності їх значень не було повторювань будь-якого фактору як по горизонталі, так і по вертикалі (табл. 3.1). Таким чином кожний водій кожного дня працює на різній відстані перевезень та на іншому автомобілі і кожного дня ці умови не повторюються для інших водіїв. Згідно зі складеним рандомізованим планом, наприклад, водій А на машині Z у понеділок має перевозити гірські породи на відстань 1,2 км, а у вівторок – на машині W на відстань 5,5 км і т.д. При цьому у кожній клітинці на перетині певного стовпчика та строки фіксується відповідний обсяг

та відстань перевезень.

Таблиця 3.1

Рандомізований план проведення спостережень
за обсягами перевезень гірських порід

Блок 1				
Водії	День тижня			
	ПН	ВТ	СР	ЧТ
A	1,2Z	5,5W	3,0X	2,8Y
B	3,0W	2,8Y	1,2Z	5,5X
C	2,8Y	1,2X	5,5W	3,0Z
D	5,5X	3,0Z	2,8Y	1,2W
Блок 2				
Водії	День тижня			
	ПН	ВТ	СР	ЧТ
A	1,8W	3,5Z	2,5X	4,2Y
B	4,2Y	1,8X	3,5W	2,5Z
C	3,5X	2,5Y	4,2Z	1,8W
D	2,5Z	4,2W	1,8Y	3,5X

Усереднення отриманих значень такого плану за регульованим фактором (відстань перевезення) дозволяє провести повну рандомізацію експерименту за всіма нерегульованими змінними (день тижня, водій, автомобіль), тим самим мінімізується випадковий вплив останніх на досліджувану ознаку.

3. Відтак, провівши спостереження, згідно зі складеним планом для кожної відстані перевезень отримаємо чотири значення обсягу перевезень (табл. 3.2), на підставі яких, відповідно, виводиться їх середнє значення.

Таблиця 3.2

Обсяги перевезень гірських порід

Відстань перевезення x_i , км	Обсяг перевезення y_i , т/зміну				Норма виробітку $u_{зм}$, т/зміну
1,2	18	19,5	18,5	20	19
1,8	17	17,5	16	16,5	16,7
2,5	15	14,5	14	15,5	14,7
2,8	14	13,5	13	12,5	13,2
3,0	13	12,5	12	13,5	12,7

3,5	11	12	11,5	10,5	11,2
4,2	10	9,5	8,5	9,5	9,4
5,5	6	6,5	5,5	6	6

4. Отримані норми виробітки є результатом збалансування умов експерименту і можуть бути використані для управління випадковими параметрами процесу транспортування через їх відповідність закономірності розподілу випадкових величин.

ПОСТАНОВКА ЗАВДАННЯ № 4

Регресійна модель. Коефіцієнти детермінації і кореляції

З використанням середньостатистичних даних, отриманих у результаті виконання завдання 3, встановіть кореляційну залежність будь-якого техніко-економічного показника діяльності організації (підприємства, виробничого підрозділу, механізму тощо) від однієї з керованих факторних перемінних (ознак), які впливають на цей показник. З цією метою виконайте наступне:

- 1) побудуйте графічну залежність;
- 2) виберіть форму рівняння зв'язку для описання вказаної залежності;
- 3) встановіть коефіцієнти рівняння кореляції та напишіть це рівняння;
- 4) оцініть тісноту зв'язку між ознаками, що корелюють, для чого розрахуйте коефіцієнт кореляції, його похибку та надійність;
- 5) сформулюйте висновок стосовно можливості використання одержаного рівняння кореляції для прогнозування показника, який розглядається у цьому завданні.

Методика вирішення

Визначається кількісна залежність між показниками, один з яких є залежною некерованою змінною, а інший таким, величина якого визначається довільно. У результаті статистичних спостережень дослідник задає числові значення незалежної змінної x і отримує відповідні значення залежної змінної y . Необхідно встановити, яким чином буде змінюватися залежна ознака за зміни значення регульованої ознаки. Якщо припускається, що зв'язок між ними не є функціональним, тобто таким, за якого відсутній вплив сторонніх чинників, а має ознаки стохастичної залежності, вважається доцільним побудувати регресійну модель, яка відображає тенденцію зміни досліджуваної ознаки під впливом її факторної величини. Якщо є можливим припустити, що на результатний показник справляє істотний вплив лише одна регульована змінна, мова йде про парну кореляцію, за якої вважається, що величина залежної змінної має стійкий причинний зв'язок лише з одним фактором та змінюється пропорційно останньому. Модель парної кореляції уможливорює використання залежності як за ретроспективного, так і перспективного аналітичного дослідження. Для кожної моделі встановлюється коефіцієнт кореляції

ції, що приймає значення від $0 \dots 1$, та відображає тісноту зв'язку між ознаками. Чим більше його значення, тим меншим є вплив випадкових факторів дослідження і тим більше можливостей є справляти управлінський вплив на досліджувану систему показників.

Методика складається з наступних етапів:

1. З використанням отриманої статистичної вибірки будується графічна залежність між результатною та факторними змінними, де кожному значенню факторної ознаки (x_i), що позначається на вісі абсцис, відповідає певне значення залежної змінної (y_i) на вісі ординат. Точка перетину відповідних значень x_i та y_i є точкою графіку функції, що описує залежність досліджуваних величин.

2. На підставі прийнятої графічної залежності та характеру кривої функції обирається форма рівняння зв'язку між величинами, що досліджується. Регресійна модель парної кореляції вміщує дві константи, за допомогою яких власне і встановлюється відповідність між змінними величинами. У моделі парної кореляції $y_i = a + bx_i$, коефіцієнти a та b є результатом впливу на досліджувану ознаку, відповідно, випадкових невстановлених чинників та керуваної факторної ознаки. Побудова моделі складається з встановлення значень коефіцієнтів регресії. Якщо для цього використовується метод найменших квадратів, та статистичну залежність нелінійного типу необхідно привести у лінійний вигляд та знайти коефіцієнти a та b вирішив систему двох рівнянь приватних похідних останніх.

3. Визначають коефіцієнти a та b рівняння регресії за наступними формулами:

$$b = \frac{\sum x_i \cdot y_i - \bar{x} \cdot \bar{y}}{\frac{\sum x_i^2}{n} - \bar{x}^2}, \quad (4.1)$$

$$a = \bar{y} - b \cdot \bar{x}, \quad (4.2)$$

де x_i, y_i – відповідно, фактичні значення досліджуваних ознак x та y ;

\bar{y}, \bar{x} – відповідно, середні значення ознак y та x ;

n – число спостережень.

4. Розраховують показники, за допомогою яких перевіряється надійність встановленої залежності ознак:

а) визначається коефіцієнт кореляції за виразом:

$$r = \frac{\sum x_i \cdot y_i - \bar{x} \cdot \bar{y}}{\delta_x \cdot \delta_y}, \quad (4.3)$$

де δ_x, δ_y – відповідно, середньоквадратичне відхилення вибірки за ознаками x та y , які досліджуються;

Середні квадратичні відхилення факторної (x) та результатної (y) ознак

встановлюють наступним чином:

$$\delta_x = \sqrt{\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right)}; \quad \delta_y = \sqrt{\left(\frac{\sum y_i^2}{n} - \bar{y}^2\right)}.$$

б) знаходиться погрішність коефіцієнта кореляції за виразом:

$$S_r = \sqrt{\frac{1-r^2}{n-2}}; \quad (4.4)$$

в) встановлюється надійність коефіцієнта кореляції:

$$M = \frac{r}{S_r}. \quad (4.5)$$

5. Формулюється висновок стосовно отриманого рівняння регресії та показників, що характеризують тісноту їх зв'язку.

Приклад виконання завдання

У результаті проведення дослідження (див. завд. 3) були отримані результати спостережень за обсягом перевезень гірської маси автомобілями на різну відстань (табл.1). Необхідно встановити залежність між обсягом виробітку автомобіля (y) та відстанню перевезень (x) і перевірити отриману залежність показників на адекватність.

Таблиця 4.1

Обсяги перевезень на різну відстань

Відстань перевезення x_i , км	Обсяг перевезення y_i , т/зміну
1,2	19
1,8	16,7
2,5	14,7
2,8	13,2
3,0	12,7
3,5	11,2
4,2	9,4
5,5	6

1. Будується графічна залежність за результатами спостережень у вигляді графіку функції $y = f(x)$, що проходить через точки перетину всіх наявних даних досліджуваних ознак (рис. 4.1):

ПОСТАНОВКА ЗАВДАННЯ № 5

Виявлення грубих помилок

1. Введіть у сукупність середньостатистичних даних, отриманих у результаті виконання завдання 4, один результат спостережень, що за величиною дуже відрізняється від загальної закономірності зміни даних, яку відображає сукупність.

2. З урахуванням цього результату (грубої помилки чи промаху), встановіть кореляційну залежність техніко-економічного показника від факторної перемінної у порядку, визначеному в завданні 4:

- а) побудуйте графічну залежність;
- б) виберіть форму рівняння зв'язку для описання вказаної залежності;
- в) встановіть коефіцієнти рівняння кореляції та напишіть це рівняння;
- г) оцініть тісноту зв'язку між ознаками, що корелюють, для чого розрахуйте коефіцієнт кореляції, його похибку та надійність;

3. Оцініть, наскільки змінилися показники тісноти зв'язку між ознаками, що досліджуються, при наявності в статистичній залежності грубих помилок, для чого отримані показники порівняйте з відповідними показниками в завданні 4 ;

4. Викладіть методику виявлення грубих помилок;

5. Здійсніть перевірку можливості вилучення результату спостережень, який дуже відрізняється від загальної закономірності зміни даних, із статистичної сукупності даних.

6. Сформулюйте висновок стосовно одержаного результату перевірки.

Методика вирішення

Дані, отримані в результаті проведення спостережень (завд. 3) повинні відображати загальну закономірність зміни результатного показника під впливом керованого фактору. Проте, не виключено, що в систему показників потрапили помилкові значення, викликані неврахованими випадковими чинниками, що дасть неправильне уявлення про існуючу залежність між досліджу-

ваними ознаками. З метою уникнення цього випадку, значення, що викликають сумнів щодо своєї достовірності, перевіряються на відповідність загальному розподілу ознак, інакше їх необхідно виключити з сукупності середньостатистичних даних. Якщо вірогідність появи певного показника є настільки малою, що практично можливою є поява ще більш відмінного значення, то перше не сприймається як достовірне і підлягає виключенню.

Методика вирішення складається з наступних етапів:

1. Дані статистичної сукупності, отримані в результаті збалансування умов експерименту (див. завд. 3), повинні бути розподілені нормально. Тому, необхідно замість певного показника, ввести інший, що порушувало би загальну закономірність зміни залежної величини за послідовної зміни керованого фактору.

2. Встановлюється кореляційна залежність між показниками, що досліджуються, за методикою, визначеною у пп. 1- 4 завд. 4 з урахуванням значення, яке суттєво відрізняється від інших.

3. Проводиться порівняння показників тісноти зв'язку сукупності даних зі зміненим значенням і вихідної сукупності та робиться висновок стосовно того, як підвищився зв'язок між ознаками для управління показниками досліджуваного процесу.

4. Викладається методика перевірки статистичної сукупності на наявність грубої помилки за наступним планом:

а) встановлюється критичне значення критерію Шовене, що відображає граничну ймовірність потрапляння результату спостережень, що сильно відхиляється від загальної залежності, у довірчий інтервал отриманої статистичної сукупності за наступною формулою:

$$P_{ш} = \frac{1}{2n}, \quad (5.1)$$

де n – число спостережень (обсяг статистичної вибірки);

б) визначають ймовірне відхилення за виразом:

$$v = \frac{R_{\epsilon}}{4}, \quad (5.2)$$

де R_{ϵ} – розмах середньостатистичної сукупності ($x_{\max} - x_{\min}$), за виключенням значення, що перевіряється;

в) розраховують показник точності даних вибірки :

$$\eta = \frac{0,477}{v}; \quad (5.3)$$

г) встановлюється розмір відхилення показника y_i , що перевіряється, від середнього значення $\overline{y_{\epsilon (n-1)}}$ сукупності, обрахованого за виключенням грубої помилки:

$$y_{\epsilon} = y_i - \overline{y_{\epsilon (n-1)}}, \quad (5.4)$$

де y_i – показник вибірки, що досліджується;

$\overline{y_{\epsilon (n-1)}}$ – середнє значення вибірки без врахування показника, що перевіряється;

д) оцінюють за табл. 5.1 вірогідність $P_{\eta y_{\epsilon}}$ потрапляння відхилення y_{ϵ} в інтервал від $+\eta y_{\epsilon}$ до $-\eta y_{\epsilon}$;

Таблиця

5.1

Вірогідність випадкових відхилень

ηy_{ϵ}	$P_{\eta y_{\epsilon}}$	ηy_{ϵ}	$P_{\eta y_{\epsilon}}$	ηy_{ϵ}	$P_{\eta y_{\epsilon}}$
0,00	0,000	0,477	0,500	0,90	0,797
0,05	0,056	0,50	0,521	0,95	0,821
0,10	0,113	0,55	0,563	1,00	0,843
0,15	0,168	0,60	0,604	1,1	0,880
0,20	0,223	0,65	0,642	1,2	0,910
0,25	0,276	0,70	0,678	1,3	0,934
0,30	0,329	0,707	0,682	1,4	0,952
0,35	0,379	0,75	0,711	1,5	0,966
0,40	0,428	0,80	0,742	2,0	0,995
0,45	0,476	0,85	0,771	∞	1,000

е) визначається вірогідність потрапляння $P_{y_{\epsilon}}$ будь - якого відхилення,

більшого за y_e , в заданий інтервал:

$$P_{y_e} = 1 - P_{ny_e}; \quad (5.5)$$

ж) порівнюють вірогідність P_{y_e} з граничним критерієм Шовене $P_{ш}$ і роблять висновок стосовно можливості виключення помилкового результату спостережень. Якщо $P_{y_e} > P_{ш}$, то показник, що перевіряється, не може бути виключеним із статистичної сукупності даних.

5. Здійснюється перевірка статистичної сукупності на наявність грубої помилки згідно з методикою, викладеною у п. 4.

6. Формулюється висновок стосовно можливості виключення значення, що перевіряється, із статистичної сукупності.

Приклад виконання завдання

1. Підставимо замість одного показника статистичної сукупності, отриманої у завд. 3, інший показник, з метою його перевірки на відповідність закономірності розподілу даних, а саме, замість значення обсягу виробітку 13,2 т/зміну (табл. 4.1), яка відповідає відстані транспортування 2,8 км, обсяг виробітку, рівний 18,0 т/зміну (табл. 5.2).

Таблиця 5.2

Обсяги перевезень на різну відстань

Відстань перевезення x_i , км	Обсяг виробітку y_i , т/зміну
1,2	19
1,8	16,7
2,5	14,7
2,8	18,0
3,0	12,7
3,5	11,2
4,2	9,4
5,5	6

2-3. Встановлюємо кореляційну залежність між ознаками, що досліджуються, та порівнюємо її показники з аналогічними показниками, отриманими у завд. 4:

а) побудуємо графік функції залежності досліджуваних ознак:

Виходячи з рис. 5.1 видно, що значення 18,0 т/зміну, яке перевіряється, порушує загальну тенденцію зміни функції. Отже, його треба перевірити на достовірність;

б) припускається, що побудований графік функції описує лінійну залежність типу $y = a + bx$;

в) визначаються коефіцієнти регресії a та b за формулами (4.1) та (4.2), приведеними у п. 3 завд. 4: $b = -3,10$, $a = 22,98$. Отже, рівняння регресії має наступний вигляд: $y = 22,98 - 3,12x$. У той же час, рівняння регресії до заміни одного результату спостережень мало вигляд: $y = 22,08 - 3,012x$. Отже, цей показник спричинив суттєве порушення зв'язку досліджуваних ознак;



Рис. 5.1. Залежність обсягу виробітку від відстані перевезення

г) розраховується коефіцієнт кореляції ознак, що досліджуються за формулою (4.3): $r = 0,94$, що свідчить про послаблення зв'язку між факторним та результативним показниками проти $r = 0,995$ за попереднього розрахунку.

Встановлюється погрішність коефіцієнту кореляції $S_r = 0,014$ та його надійність $M = 6,69$, відповідно, за формулами (4.4) та (4.5). Відтак, погрішність коефіцієнту кореляції проти $S_r = 0,035$ зросла, а надійність проти $M = 28,17$ – зменшилась, що вказує на необхідність перевірки зазначеного результату спостережень на його адекватність встановленої раніше закономірності розподілу даних.

4. Описується, яким чином необхідно перевірити статистичну сукупність на наявність грубої помилки.

5. Здійснюється перевірка статистичної сукупності на наявність помилкового значення, згідно з методикою, викладеною у вище наведеному п. 4:

а) розраховується критичне значення критерію Шовене за формулою (5.1) для наявної сукупності усереднених даних:

$$P_{ш} = 1/(2 \cdot 8) = 0,0625.$$

б) визначається ймовірне відхилення сукупності даних за формулою (5.2):

$$v = 3,25.$$

в) обчислюється показник точності даних вибірки за формулою (5.3):

$$\eta = 0,137.$$

г) встановлюється за формулою (5.4) величина відхилення показника, що перевіряється, від середнього значення статистичної сукупності: $y_e = 12,81$;

д) встановлюється за табл. 5.1 вірогідність $P_{\eta y_e}$ потрапляння відхилення у інтервал від $+0,137 \cdot 12,81$ до $-0,137 \cdot 12,81$:

$$P_{\eta y_e} = 0,997;$$

е) визначається вірогідність непотрапляння P_{y_e} відхилення у заданий ін-

тервал за формулою (5.5):

$$P_{y_0} = 1 - 0,997 = 0,003.$$

6. Висновок: Так-як вірогідність непотрапляння заданого відхилення ознаки 0,003 є меншою критичного значення Шовене, яке для даної сукупності обсягів перевезення дорівнює 0,0625, то показник обсягу перевезення 18,0 т/зміну повинен бути виключеним із статистичної вибірки.



Рис. 4.1. Залежність обсягу виробітку від відстані перевезення

2. Виходячи з рис. 4.1 є очевидним, що між показниками існує лінійна залежність типу $y = a + bx$: поступове збільшення відстані перевезення обумовлює відповідне зменшення обсягу виробітку.

3. Визначаються коефіцієнти регресії a та b за формулами (4.1) і (4.2). Отримано: $b = - 3,012$, $a = 22,08$. Отже, рівняння регресії має наступний вигляд: $y = 22,08 - 3,012 x$.

4. Розраховується коефіцієнт кореляції ознак, що досліджуються, за формулою (4.3). Для встановленого рівняння регресії $r = 0,995$, що свідчить про дуже тісний зв'язок між факторним та результативним показниками.

Встановлюється погрішність коефіцієнту кореляції та його надійність за виразами відповідно (4.4) та (4.5). Ці характеристики дорівнюють: $S_r = 0,037$, $M = 26,89$.

5. Згідно з отриманим рівнянням збільшення відстані перевезень на один кілометр призводить до зменшення обсягу виробітку на 3 т. Отриманий коефіцієнт кореляції, а також його погрішність та надійність свідчать про стійкий зв'язок між корельованими ознаками, що робить модель парної кореляції придатною для використання у практичних розрахунках норми виробітку.

