Семінар 2. Розв’язок задач

Дискретна випадкова величина. Нехай $(\Omega, \Im, P)$ — ймовірнісний простір. Дискретною випадковою величиною називається функція $\xi(w)$ на Ω, яка набуває скінченне або зліченне число значень x_1, x_2..., x_n, ... і є вимірною відносно σ-алгебри $\Im$.

Математичне сподівання випадкової величини. Нехай $\xi(w)$ — дискретна випадкова величина, яка набуває значень x_i, з імовірностями p_i (і = 1, 2,...). Припустимо, що ряд $\sum_i |x_i| p_i < +\infty$ збігається. Тоді математичним сподіванням випадкової величини $\xi(w)$ називається сума ряду $M\,\xi(w) = \sum_i x_i\, p_i$.

Властивості математичного сподівання:

1. MC=C, де C — константа.
2. Якщо $\xi \geq 0$, то $M\xi \geq 0$.
3. $M(C\xi) = CM\xi$.
4. $M(\xi \pm C) = M\xi \pm C$.
5. $M(\xi \pm \eta) = M\xi \pm M\eta$.
6. Якщо ξ та η — незалежні випадкові величини, то $M(\xi\eta) = M\xi\, M\eta$.
7. $Mf(\xi) = \sum_i f(x_i) p_i$.

Дисперсія випадкової величини $\xi(w)$ визначається рівністю $D\xi = M(\xi - M\xi)^2 = M\xi^2 - (M\xi)^2 = \sum_i (x_i - M\xi)^2 p_i$.

$\sigma = \sqrt{D\xi}$ – **середньоквадратичне відхилення випадкової величини,** або **ризик**.

Властивості дисперсії:

1. $D\xi = 0 \Leftrightarrow \xi = C$ – константа.
2. $D\xi \geq 0$.
3. $D(C\xi) = C^2 D\xi$.
4. $D(\xi \pm C) = D\xi$.
5. Якщо ξ та η незалежні випадкові величини, то $D(\xi \pm \eta) = D\xi + D\eta$.

Коефіцієнт коваріації випадкових величин $cov(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta)$.

Коефіцієнт кореляції. Коефіцієнтом кореляції випадкових величин називається

$$\rho(\xi,\eta) = \frac{M(\xi - M\xi)(\eta - M\eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{cov(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}},\ D\xi > 0,\ D\eta > 0.$$

Справджуються такі твердження:

а) $|\rho| \le 1$;

б) якщо ξ та η незалежні, то $\rho(\xi, \eta) = 0$;

в) якщо $|\rho| = 1$, то з імовірністю одиниця $\eta = a + b\xi$, де a і b — деякі сталі, $a \neq 0$.

Незміщеною, сильно спроможною та асимптотично нормальною оцінкою математичного сподівання є вибіркове середнє $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$. Через $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$ будемо позначати реалізацію цієї оцінки. Вибіркове середнє для дискретного статистичного ряду підраховується за формулою $\bar{x} = \frac{1}{n}\sum_{i=1}^{s}y_i m_i$, $n = \sum_{i=1}^{s}m_i$. Вибіркове середнє для інтервального статистичного ряду підраховується за формулою $\bar{x} = \frac{1}{n}\sum_{i=1}^{s}y_i^* m_i$, де $y_i^* = \frac{y_i + y_{i+1}}{2}$ – середина інтервалу $[y_i, y_{i+1})$.

$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2$ — незміщена оцінка дисперсії σ^2.

$\bar{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2$ — зміщена оцінка σ^2.

Якщо математичне сподівання a відоме, то незміщеною, сильно спроможною і асимптотично нормальною оцінкою дисперсії є оцінка $S^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - a)^2$.

Приклад розв'язання задачі

Задача 1. Випадкова величина ξ має розподіл:

x_i	-1	0	1	2
p_i	0,2	0,1	0,3	0,4

Знайти математичне сподівання і дисперсію випадкової величини $\eta = 2^{\xi}$.

Розв'язок. Будуємо розподіл випадкової величини η:

y_i	0,5	1	2	4
p_i	0,2	0,1	0,3	0,4

Математичне сподівання $M\eta$=0,5·0,2+1·0,1+2·0,3+4·0,4=2,4.

Дисперсія $D\eta$=1,9^2·0,2+1,4^2·0,1+0,4^2·0,3+1,6^2·0,4=1,99.

Приклад розв'язання задачі

На основі статистичних даних доходу підприємства (у млн. грн.) y та кількості працюючих (у тис. чол.) x

y	x
10,8	2,53
11,9	3,54
12,4	3,84
13,2	3,84
14,1	4,22
15,2	4,81
16,0	6,53
17,4	5,82
18,6	6,43
19,4	7,73
20,5	8,19
21,3	7,65
22,5	9,31
23,7	9,26
25,0	9,86

Знайти оцінки параметрів лінійної регресії $y = \alpha + \beta x + \varepsilon$.

Розв'язок.

Заповнюємо таблицю:

y	x	x^2	y^2	xy
10,8	2,53	6,40	116,64	27,32
11,9	3,54	12,53	141,61	42,13
12,4	3,84	14,75	153,76	47,62
13,2	3,84	14,75	174,24	50,69
14,1	4,22	17,81	198,81	59,50
15,2	4,81	23,14	231,04	73,11

	16,0	6,53	42,64	256,00	104,48
	17,4	5,82	33,87	302,76	101,27
	18,6	6,43	41,34	345,96	119,60
	19,4	7,73	59,75	376,36	149,96
	20,5	8,19	67,08	420,25	167,90
	21,3	7,65	58,52	453,69	162,95
	22,5	9,31	86,68	506,25	209,48
	23,7	9,26	85,75	561,69	219,46
	25,0	9,86	97,22	625,00	246,50
Сума	262,00	93,56	662,22	4864,06	1781,95

Параметри регресії знаходяться за формулами

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2};\ \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}.$$

Тоді

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{15 \cdot 1781,95 - 93,56 \cdot 262}{15 \cdot 662,22 - \left(93,56\right)^2} = 1,8787,$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} = \frac{\sum y_i}{n} - \hat{\beta}\frac{\sum x_i}{n} = \frac{262}{15} - 1,8787\frac{93,56}{15} = 5,7486.$$

Таким чином вибіркова регресійна функція записується у такому вигляді:

$$\hat{y} = 5,7486 + 1,8787x.$$

При збільшенні кількості працюючих на 1000 чоловік доход підприємства зростає на 1,8787 млн. грн.

$\hat{y}_i$	$\left(y_i - \hat{y}_i\right)^2$	$\left(\hat{y}_i - \overline{y}\right)^2$	$\left(y_i - \overline{y}\right)^2$
10,502	0,089	48,511	44,444
12,399	0,249	25,679	30,988
12,963	0,317	20,285	25,671

12,963	0,056	20,285	18,204
13,677	0,179	14,364	11,334
14,785	0,172	7,191	5,138
18,017	4,066	0,302	2,151
16,683	0,515	0,615	0,004
17,829	0,595	0,131	1,284
20,271	0,759	7,864	3,738
21,135	0,403	13,458	9,201
20,121	1,391	7,044	14,694
23,239	0,547	33,323	25,334
23,145	0,308	32,248	38,854
24,273	0,529	46,320	56,751
Сума	10,175	277,618	287,793

Таким чином, $RSS = 10{,}175$, $ESS = 277{,}618$, $TSS = 287{,}793$.

Коефіцієнт детермінації моделі становить $R^2 = \dfrac{ESS}{TSS} = \dfrac{277{,}618}{287{,}793} = 0{,}965$, що свідчить про високу степінь залежності між змінними.

Розв'язати задачі

1. Двічі кидають монету. Описати простір елементарних подій Ω. Нехай $\xi(w)$ — число випадань герба. Знайти розподіл випадкової величини ξ, математичне сподівання $M\xi$ та дисперсію $D\xi$.
2. Двічі підкидають гральний кубик. Описати простір елементарних подій Ω. Нехай $\xi(w)$ — сума очок, які випали. Знайти розподіл випадкової величини ξ, $M\xi$.
3. Математичне сподівання та дисперсія випадкової величини ξ дорівнюють відповідно 2 та 10. Знайти математичне сподівання та дисперсію величини $2\xi+5$.
4. Знайти середнє квадратичне відхилення випадкової величини, поданої розподілом:

ξ	3	5	7	9
P	0,4	0,3	0,2	0,1

5. Обчислити незміщені оцінки математичного сподівання та дисперсії, вибіркову медіану та моду для вибірки:

y_i	12	14	16	18	20	22
m_i	5	15	50	16	10	4

6. Обчислити вибіркове середнє і дисперсію, моду, медіану для вибірки:

Інтервал	[2, 4)	[4, 6)	[6, 10)	[10, 16)	[16, 20)
m_i	2	8	35	40	15

7. Знайти перетворення даних, яке зводить дану модель в лінійну. Визначити, яким чином потрібно включити збурення до моделі:

a) $y = \dfrac{1}{\alpha + \beta \cdot e^x}$;

b) $y = \dfrac{e^x}{\alpha + \beta \cdot e^x}$;

c) $y = 1 + \alpha\left(1 - x^\beta\right)$;

d) $y^2 = \alpha e^{\beta x}$.

8. Знайти невідомий коефіцієнт моделі $y_t = \alpha + \varepsilon_t, t = \overline{1,n}$. Чому дорівнює коефіцієнт детермінації моделі? Знайти дисперсію моделі. Покажіть, що статистика $\dfrac{\hat{\alpha} - \alpha}{s.e.\left(\hat{\alpha}\right)}$ має t_{n-1}-розподіл.

9. Знайти невідомий коефіцієнт моделі $y_t = \beta x_t + \varepsilon_t, t = \overline{1,n}$. Знайти коефіцієнт детермінації моделі.

10. Спостереження 16 пар $(x,\ y)$ дали результати: $\sum y^2 = 526$, $\sum x^2 = 657$, $\sum xy = 492$, $\sum y = 64$, $\sum x = 96$. Оцінити регресію $y_t = \alpha + \beta x_t + \varepsilon_t$. Знайдіть коефіцієнт детермінації. Підрахуйте найкращий незміщений прогноз на наступний період, якщо x_{n+1}=4.

11. Для моделі $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, $t = \overline{1,20}$, відомо, що $\sum y_t = 15$, $\sum x_t = 41$, $\sum y_t^2 = 29$, $\sum x_t^2 = 214$, $\sum x_t y_t = 31$. Відомо, що $x_{21} = 10$. Обчислити

найкращий лінійний незміщений прогноз величини $\hat{y}_{21}$. Оцінити стандартну похибку прогнозу.

12. Для спостережень

Y	5	13	12	17	12	22	15	22	33	35
X	90	25	42	50	36	35	12	60	25	32

обчисліть величину коефіцієнта детермінації R^2 в регресії Y_t на X_t при відсутності вільного члена. Проаналізуйте отриману регресію. Підрахуйте оцінку прогнозу при $x_{n+1} = 23$.

13. Для спостережень

Y	5	11	12	17	10	22	15	27	30	35
X	70	65	35	60	46	35	42	30	25	32

обчисліть величину коефіцієнта детермінації R^2 в регресії Y_t на X_t при наявності вільного члена. Проаналізуйте отриману регресію. Оцініть прогноз при $x_{n+1} = 52$. Перевірте адекватність моделі, $\alpha = 0{,}01$.

14. Оцінити модель простої лінійної регресії $y = \alpha + \beta x + \varepsilon$ за даними

Y	320	325	298	307	301	302	290
X	45	43	42	40	38	39	35

та перевірте її на адекватність, $\alpha = 0{,}05$.

15. Монополіст максимізує прибуток за наявності наступної функції попиту $Q = \alpha + \beta P + \varepsilon$. У минулому спостерігалася така залежність між цінами та рівнем продаж:

Q	3	3	7	6	10	15	16	13	9	15
P	18	16	17	12	15	15	4	13	11	6

Визначити максимальний доход монополіста. Знайти ціну, яку запропонує монополіст. Побудувати 90% надійний інтервал для випуску, що максимізує прибуток.