Олимпиада студентов и выпускников «Высшая лига» – 2021 г.

Критерии оценивания и решения олимпиадных заданий заключительного этапа по направлению «230. Прикладная математика»

Инвариантная часть

Задание 1 (11 баллов).

По 21 наблюдению из нормального распределения с неизвестными математическим ожиданием и дисперсией был оценен 90% доверительный интервал для матожидания: [0.825; 11.175]

1) Можно ли на основе этих данных утверждать, что гипотеза о равенстве математического ожидания рассматриваемой случайной величины нулю H0: $\mu = 0$ отвергается на 10% уровне значимости?
2) Найдите выборочное среднее и выборочную оценку дисперсии
3) Проверьте гипотезу H0: $\mu = 2$ на 5% уровне значимости

Решение:

1. Можно, гипотеза отвергается на 10% уровне значимости, так как $0 \notin [0.8254; 11.175]$, где [0.8254; 11.175] – 90% уровень значимости.
2. Выборочное среднее = 6, выборочная оценка дисперсии ≈ 13,745.
3. Гипотеза H0 принимается: $\mu = 2$ на 5% уровне значимости.

Максимальный балл выставлялся за наличие трёх правильных ответов, сопровождённых ходом решения.

Задание 2 (11 баллов).

Рассмотрите фрагмент данных проведённого в России исследования об использовании информационно-коммуникационных технологий. В ходе психологическогоопроса респонденты ответили на два вопроса, в частности.

Вопрос 1: «Сколько часов в день Вы проводите в интернете – по работе или с личными целями?».

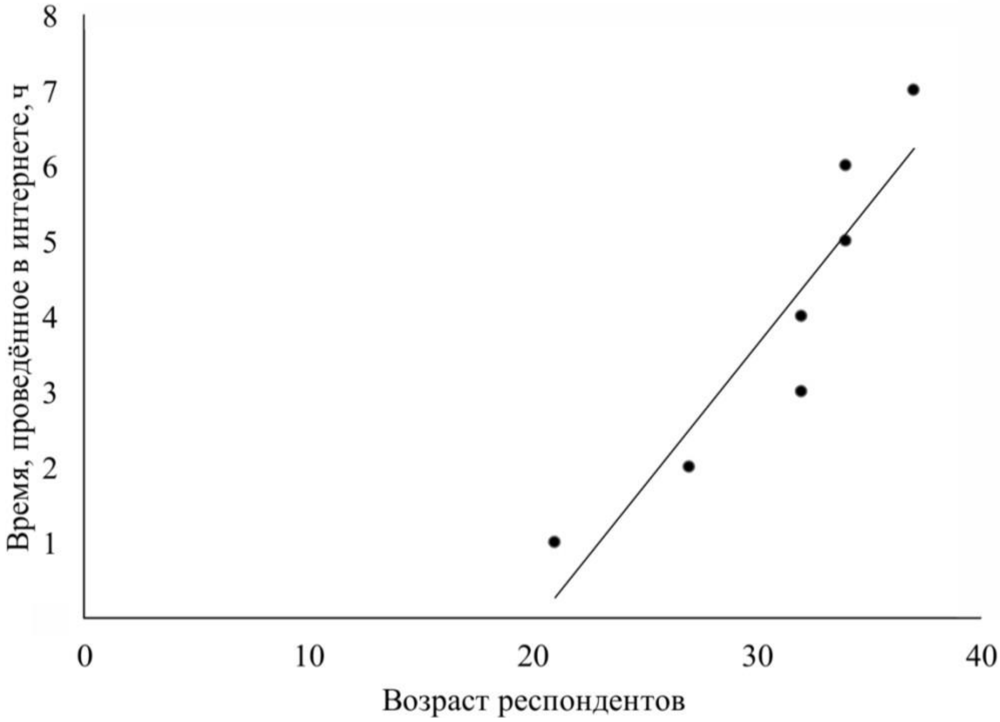
Вопрос 2: «Укажите, пожалуйста, ваш возраст».

Можно ли считать два этих показателя линейно коррелированными? Рассчитайте коэффициенты корреляции Пирсона. Интерпретируйте полученные ответы. Осуществите проверку статистической значимости коэффициента, приняв уровень доверительной вероятности равным 0,95. Интерпретируйте полученный ответ. Ответьте на вопрос о том, есть ли справедливые основания утверждать на уровне доверительной вероятности 0,95, что в среднем участники исследования проводят в интернете около 4 часов, используя методику расчета доверительного интервала.

Респондент	Время, проведённое в интернете, ч.	Возраст
1. Алексей	1	21
2. Мария	3	32
3. Екатерина	2	27
4. Дмитрий	5	34
5. Николай	4	32
6. Александр	6	34
7. Ольга	7	37

Решение.

Часть 1. Для предварительной оценки характера линейной взаимосвязи построим диаграмму рассеяния на основе выше представленных признаков – время, проведённое в интернете, и возраст респондента – и схематично нарисуем линию тренда.



Далее найдём среднее арифметическое:

$\bar{x} = \frac{1+3+2+5+4+6+7}{7} = 4$, где $\bar{x}$ – среднее время, проведённое в интернете.

$\bar{y} = \frac{21+32+27+34+32+34+37}{7} = 31$, где $\bar{y}$ – средний возраст участников исследования.

Теперь определим коэффициент линейной корреляции. Вычислим основные компоненты формулы для определения коэффициента линейной корреляции, построиврасчётную таблицу.

x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})*(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	21	-3	-10	30	9	100
3	32	-1	1	-1	1	1
2	27	-2	-4	8	4	16
5	34	1	3	3	1	9
4	32	0	1	0	0	1
6	34	2	3	6	4	9
7	37	3	6	18	9	36
$\bar{x}=4$	$\bar{y}=31$	—	—	$\sum_{i=1}^{n=7}(x_i-\bar{x})*(y_i-\bar{y})=64$	$\sum_{i=1}^{n=7}(x_i-\bar{x})^2=28$	$\sum_{i=1}^{n=7}(y_i-\bar{y})^2=172$

И подставим в формулу полученные значения по основным компонентам:

$$R_{\text{Пирсон}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})*(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2*\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{64}{\sqrt{28}\sqrt{172}} = \frac{64}{5{,}29*13{,}11} \approx 0{,}92.$$

Положительное значение коэффициента линейной корреляции указывает на наличие*прямой* зависимости между временем, проведённым респондентом в интернете, и его возрастом. Для данного набора наблюдений можно считать рассматриваемые признаки линейно коррелированными на очень высоком уровне интенсивности ($R \approx 0{,}92$).

Часть 2. Проверим статистическую значимость полученного на выборке коэффициента линейной корреляции. Приступим к проверке нулевой гипотезы (линейная корреляция на самом деле отсутствует, то есть равна нулю), предварительно задав альтернативную гипотезу (линейная корреляция носит положительный характер, то есть больше нуля). Следовательно, критическая область будет правосторонней, фиксированныйуровень значимости $\alpha = 0{,}05$ (по условию задачи).

Выпишем статистический t-критерий как функцию от имеющихся данных $t = R * \sqrt{\frac{n-2}{1-R^2}}$. Найдём значение критерия $t_{\text{наблюдаемое}} = R * \sqrt{\frac{n-2}{1-R^2}} = +0{,}92 * \sqrt{\frac{7-2}{1-0{,}92^2}} \approx 5{,}25$

Укажем критическую область, воспользовавшись таблицей критических точек для t-распределения Стьюдента (для $\alpha = 0{,}05$ и $df = 7 - 2 = 5$). По таблице $t_{\text{критическое}} = +2{,}015$. Критическая область является интервалом $(+2{,}015;\ +\infty)$. Наблюдаемое значение (+5,25) входит в критическую область и не принадлежит области подтверждения нуль- гипотезы. При имеющихся данных нулевая гипотеза о равенстве нулю истинного значениякоэффициента линейной корреляции отвергается.

Значит, выборочная оценка коэффициента корреляции *статистически значима*. Обобщая результаты на генеральную совокупность, сформулируем содержательный вывод: на уровне доверительной вероятности 0,95 имеющиеся данные выборочного исследования *позволяют* считать, что такие показатели, как возраст респондентов и время, проведённое ими в интернете, *линейно коррелированы (связаны)*. Обнаруженная корреляция +0,92 справедлива можно считать статистической закономерностью для генеральной совокупности.

Ответ. (1) Положительное значение коэффициента линейной корреляции указываетна наличие прямой зависимости между временем, проведённым респондентом в интернете,и его возрастом ($R \approx 0{,}92$). (2) $t_{\text{наблюдаемое}} = 5{,}25 > t_{\text{критическое}} = 2{,}015$. Следовательно, наблюдаемое значение входит в критическую область $(+2{,}015;\ +\infty)$ и не принадлежит области подтверждения нуль-гипотезы на уровне значимости равном α =0,05. (3) При имеющихся данных проведенное исследование с вероятностью 0,95 можно утверждать, чтоистинное среднее значение количества часов, проведенных участниками исследования в интернете, достоверно содержится в границах интервала от 3,92 до 4,08 часов.

Задание 3 (11 баллов).

В ходе небольшого исследования, посвящённого теме *бюджета времени* и выборка которого случайна, в 2012 году приняли участие 4 девушки, студентки второго курса (которые в течение определённого периода заполняли специальные дневники, фиксируя информацию о том, скольком времени они выделяют на те или иные виды деятельности). Так, было зафиксировано, сколько обычно часов в неделю они расходуют *на отдых и развлечения*. Спустя год, в 2013 году исследование повторили (с тем же составом участников и по той же схеме). Были получены следующие результаты:

(в ячейках указаны данные о еженедельных расходах времени на отдых и развлечения)

	Катерина	**Ольга**	**Алла**	**Луиза**
Данные 2012 года (учились на 2 курсе)	30 часов	37 часов	45 часов	34 часов
Данные 2013 года (перешли на 3 курс)	21 часов	35 часов	40 часов	34 часов

Можно ли сказать (при имеющихся данных), что с возрастом (по мере перехода на более старший курс обучения в университете) девушки склонны в меньшей степени тратить время на отдых и развлечения? Проверку соответствующей статистической гипотезы реализуйте на уровне доверительной вероятности 0,95 против *направленной* альтернативной гипотезы.

КРИТЕРИИ ОЦЕНИВАНИЯ

Применен метод, соответствующий маленькому количеству наблюдений (решение сведено к случаю Бернулли или применен непараметрический метод или перед применением параметрического метода данные проверены на нормальность/указана необходимость проверки на нормальность), без ошибок - 11 баллов, с незначительными ошибками - 9-10 баллов	9 - 11 баллов
Применен параметрический метод без проверки данных на нормальность с учетом степеней свободы и несмещенной оценкой дисперсии без ошибок - 8 баллов, без поправки на степени свободы или несмещенную оценку дисперсии, но без других ошибок - 7-6 баллов, с ошибками и без поправок на размер выборки - 5 баллов	8-5 балоов
Заход на верный или подходящий метод, но без полноценных расчетов ИЛИ неподходящий метод с расчетами и выводом	3-4 балла
Сделано верное предположение по поводу выбора метода/захода, но расчетов нет	1-2 балла
Нет расчетов и верного предположения относительно применяемого метода	0 баллов

$$\bar{T} = \frac{n(n+1)}{4}$$

$$SE_{\bar{T}} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$z = \frac{X - \bar{X}}{s} = \frac{T - \bar{T}}{SE_{\bar{T}}}$$

T_{mean}	3
SE	1,870828693
$T_{\text{наименьшая из сумм}}$	0
n (без нулей)	3
z	-1,603567451
z крит	-1,64
Вывод	нулевая гипотеза принимается Различий нет

Задание 4 (11 баллов).

Привести пример матриц A и B таких, что $A \cdot B = I$, но $B \cdot A \neq I$, где I – единичная матрица.

Решение: В качестве примера можно рассмотреть матрицы

$$A = \begin{pmatrix} 1 & 0 \end{pmatrix} \text{ и } B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

В самом деле,

$$\begin{aligned} A \cdot B &= \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} = I_{1\times 1} \\ B \cdot A &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq I_{2\times 2} \end{aligned}$$

Критерии оценивания:

11 баллов : приведены матрицы A и B, удовлетворяющие условию задачи;

5 баллов : показано, что не существует квадратных матрицы A и B, удовлетворяющих условию.

Вариативная часть

Трек «Математические методы анализа в экономике»

Задание 5 (7 баллов).
Математический анализ №1

Evaluate the following limit: $\lim\limits_{x\to+\infty} sin\left[x\left(\ln\left(1+\frac{x}{2}\right)-\ln\frac{x}{2}\right)\right]$.

Solution

First consider the function $x\left(\ln\left(1+\frac{x}{2}\right)-\ln\frac{x}{2}\right)$ and find its limit as x tends to infinity. Note that $x\left(\ln\left(1+\frac{x}{2}\right)-\ln\frac{x}{2}\right) = x\ \ln\left(1+\frac{2}{x}\right)$ **(1.5 points)**.

By the Taylor equivalence $\lim\limits_{x\to+\infty} x\ \ln\left(1+\frac{2}{x}\right) = \lim\limits_{x\to+\infty} x\ \left(\frac{2}{x}+o\left(\frac{2}{x}\right)\right)$ **(2 points)**.

$\lim\limits_{x\to+\infty} x\ \left(\frac{2}{x}+o\left(\frac{2}{x}\right)\right) = 2 + \lim\limits_{x\to+\infty}\left(x\cdot o\left(\frac{2}{x}\right)\right) = 2$ **(2 points)**.

As sin is a continuous function, $\lim\limits_{x\to+\infty} sin\left[x\left(\ln\left(1+\frac{x}{2}\right)-\ln\frac{x}{2}\right)\right] = \sin\left(\lim\limits_{x\to+\infty} x\left(\ln\left(1+\frac{x}{2}\right)-\ln\frac{x}{2}\right)\right) = sin(2)$. **(2 points)**.

Задание 6 (7 баллов).
Математический анализ №2

Let

$$\xi = \sum_{k=0}^{\infty}\frac{1}{2^{k^2}} = 1+\frac{1}{2}+\frac{1}{16}+\cdots$$

Find the limit

$$\lim_{n\to\infty} 4^n \sin\left(\pi\xi 2^{n^2}\right)$$

or prove that this limit does not exist.

Solution

1) Обозначим

$$r_n = \sum_{k=n}^{\infty} \frac{1}{2^{k^2}} = \frac{1}{2^{n^2}} + \frac{1}{2^{(n+1)^2}} + \cdots$$

Тогда

$$\sin\left(\pi\xi 2^{n^2}\right) = \sin\left(\pi\left(1 + \frac{1}{2} + \cdots + \frac{1}{2^{(n-1)^2}} + \frac{1}{2^{n^2}} + r_{n+1}\right)2^{n^2}\right)$$
$$= \sin\left(\pi\left(1 + \frac{1}{2} + \cdots + \frac{1}{2^{(n-1)^2}}\right)2^{n^2} + \pi + \pi r_{n+1}2^{n^2}\right)$$

Первое слагаемое отбрасываем, т. к. оно кратно 2π, затем пользуемся формулой $\sin(\pi + x) = -\sin(x)$. Получаем:

$$\sin\left(\pi\xi 2^{n^2}\right) = \sin\left(\pi + \pi r_{n+1}2^{n^2}\right) = -\sin\left(\pi r_{n+1}2^{n^2}\right)$$

(2,5 point).

2) Оценим $r_{n+1}2^{n^2}$:

$$r_{n+1}2^{n^2} = \left(\frac{1}{2^{(n+1)^2}} + \sum_{k=2}^{\infty}\frac{1}{2^{(n+k)^2}}\right)2^{n^2} = \frac{2^{n^2}}{2^{(n+1)^2}} + \frac{2^{n^2}}{2^{(n+2)^2}}\sum_{k=2}^{\infty}\frac{2^{(n+2)^2}}{2^{(n+k)^2}}$$
$$= \frac{1}{2^{2n+1}} + \frac{1}{2^{4n+4}}\sum_{k=2}^{\infty}\frac{1}{2^{(k-2)(2n+k+2)}}$$

Ряд $\sum_{k=2}^{\infty}\frac{1}{2^{(k-2)(2n+k+2)}}$ сходится, а $\frac{1}{2^{4n+4}} = o\left(\frac{1}{2^{2n+1}}\right)$ при $n \to \infty$.
Поэтому при $n \to \infty$.

$$r_{n+1}2^{n^2} = \frac{1}{2^{2n+1}} + o\left(\frac{1}{2^{2n}}\right)$$

(3 point).

3) Следовательно, при $n \to \infty$

$$\sin\left(\pi\xi 2^{n^2}\right) = -\sin\left(\pi r_{n+1}2^{n^2}\right) = -\frac{\pi}{2^{2n+1}} + o\left(\pi\left(\frac{1}{2^{2n+1}} + o\left(\frac{1}{2^{2n}}\right)\right)\right)$$
$$= -\frac{\pi}{2^{2n+1}} + o\left(\frac{1}{2^{2n}}\right)$$

(1point).

4) Окончательно,

$$\lim_{n\to\infty} 4^n \sin\left(\pi\xi 2^{n^2}\right) = \lim_{n\to\infty} 2^{2n}\left(-\frac{\pi}{2^{2n+1}} + o\left(\frac{1}{2^{2n}}\right)\right) = -\frac{\pi}{2}$$

(1 point).

Задание 7 (7 баллов).

Линейная алгебра №1

For each integer $n \geq 2$, let A_n be the $n \times n$ matrix (a_{ij}) satisfying

$$a_{ij} = \begin{cases} 2-n & \text{if} \quad i=j \\ 1 & \text{otherwise} \end{cases}$$

i.e.

$$A_n = \begin{pmatrix} 2-n & 1 & \dots & 1 \\ 1 & 2-n & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 2-n \end{pmatrix}.$$

Prove that the matrix A_n is invertible and find A_n^{-1}.

Solution

1) Заметим, что i-ый столбец $\boldsymbol{c}_i$ матрицы A^{-1} есть решение уравнения $A\boldsymbol{c}_i = \boldsymbol{b}_i$, где $\boldsymbol{b}_i$ есть вектор-столбец, содержащий единицу на i-ом месте и нули в остальных местах. Если все такие уравнения имеют решение, то матрица A обратима. **(2 point)**.

2) Решим уравнение $A_n\boldsymbol{c}_i = \boldsymbol{b}_i$. Пусть $\boldsymbol{c}_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Тогда

$$\begin{cases} (2-n)x_1 + x_2 + \dots + x_i + \dots + x_n = 0 \\ x_1 + (2-n)x_2 + \dots + x_i + \dots + x_n = 0 \\ \dots \\ x_1 + x_2 + \dots + (2-n)x_i + \dots + x_n = 1 \\ \dots \\ x_1 + x_2 + \dots + x_i + \dots + (2-n)x_n = 0 \end{cases}$$

Сложим все уравнения. Получим $x_1 + x_2 + \dots + x_n = 1$. **(2 point)**.

Используя это выражение, преобразуем каждое из уравнений:

$$\begin{cases} (2-n)x_1 + 1 - x_1 = 0 \\ (2-n)x_2 + 1 - x_2 = 0 \\ \dots \\ (2-n)x_i + 1 - x_i = 1 \\ \dots \\ (2-n)x_n + 1 - x_n = 0 \end{cases}$$

Следовательно, $x_1 = x_2 = \dots = x_{i-1} = x_{i+1} = \dots = x_n = \frac{1}{n-1}$, $x_i = 0$. **(2 point)**.

Проверяем, что это действительно решение. **(1 point)**.

Окончательно, $A_n^{-1} = \frac{1}{n-1}\begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 0 \end{pmatrix}$ **(0,5 point)**.

Если задача решена иным способом (например, ответ угадан и обоснован), баллы начисляются индивидуально.

Задание 8 (7 баллов).
<u>Оптимизация</u>
Analyze the presence and nature of a conditional extremum:

A) Function. $F(x,y) = x^2 + y^2 + 2 - 2(x+y)$
B) Constraint. $G(x,y) = \sqrt[2]{y^2 + x^2 + 4(1-x) - 2y(x-2)} = 4$. Constraint contains the point $(x,y) = (6,2)$.

Solution

Путем простых преобразований задача сводится к задаче

A) $F(x,y) = (x-1)^2 + (y+1)^2$
B) $G(x,y) = \sqrt[2]{(y-x+2)^2} = 4$ или $|y-x+2| = 2$.

Условие порождает два случая:

I) $y - x + 2 = 2$, однако, оно не содержит точку $(x,y) = (6,2)$
II) $y - x + 2 = -2$, т.е. $y = x - 4$, которое содержит точку $(x,y) = (6,2)$

Таким образом надо найти исследовать наличие условного экстремума в задаче:

C) $F(x,y) = (x-1)^2 + (y+1)^2$
D) $y = x - 4$.

Здесь нет смысла прибегать к использованию функции Лагранжа. Достаточно подставить в уравнение функции ограничение. В результате получим $F(x,y) = (x-1)^2 + (x-3)^2$. Несложно определить, что точка $(x,y) = (2,-2)$ является точкой минимума.

Критерии оценивания:

1. Приведение задачи к виду C-D **3 балла**
2. Правильное нахождение экстремума в задаче C-D **3 балла**
3. Правильное исследование характера экстремума в задаче C-D **1.5 балла**

Задание 9 (7 баллов).
<u>Диф. уравнения №1</u>
Solve the following differential equation: $y\,dx = (x^2y + x)\,dy$.
Solution

1) Note that $\frac{x\,dy - y\,dx}{x^2} = d(y/x)$. **(1 point)**.
2) Dividing the differential equation by x^2, we get $\frac{-y\,dx + x\,dy}{x^2} + y\,dy = 0$, hence $d(y/x) + y\,dy = 0$. **(2 points)**.
3) Make one more transformation of differential equation: $d(y/x) + \frac{1}{2}d(y^2) = 0$, which is equivalent to $d\left(\frac{y}{x} + \frac{1}{2}y^2\right) = 0$. **(2 points)**.
4) Integrating this expression, we get the solution of differential equation $\frac{y}{x} + \frac{1}{2}y^2 = C$, where $C \in \mathbb{R}$. **(2 points).**

5) Note that we divided the equation by x^2, so we need to check if $x = 0$ is a solution of differential equation. Substituting $x = 0$ into differential equation we find it is indeed a solution. **(0.5 points)**.

Задание 10 (7 баллов).

<u>Теория вероятностей</u>

N points are selected randomly and independently in the segment $[0, 1]$ of the axis Ox of the coordinate plane Oxy (the distance from each of them to the origin is a uniformly distributed random variable). An ant is placed at each of the selected points. At the zero moment of time, each ant begins to move at a constant speed 1 in the positive direction of the axis Oy (perpendicular to Ox). When an ant reaches the straight line $x + y = 1$, it stops and no longer moves. Let the random variable Y be the moment when the last ant stops moving.

(a) Find the cumulative distribution function of Y.

(b) Find the expected value of Y.

Solution

1) Пусть X_i есть случайная величина «абсцисса точки старта i-го муравья», а T_i есть случайная величина «момент остановки i-го муравья» (муравьи занумерованы произвольно). Длина пути i-го муравья до прямой $x + y = 1$ есть случайная величина $1 - X_i$. Отсюда $T_i = \frac{1-X_i}{1} = 1 - X_i$. Поэтому для любого $t \in [0,1]$ выполнено $\mathrm{P}(T_i \le t) = \mathrm{P}(1 - X_i \le t) = \mathrm{P}(X_i \ge 1 - t) = t$. **(3 point)**

2) Заметим, что $Y = \max\{X_1, X_2, X_3, \dots, X_N\}$. **(1 points)**

3) Для любого $t \in [0,1]$ вычислим вероятность

$$\mathrm{P}(Y \le t) = \mathrm{P}\big((X_1 \le t)(X_2 \le t) \dots (X_N \le t)\big) = \mathrm{P}(X_1 \le t)\mathrm{P}(X_2 \le t)\dots\mathrm{P}(X_N \le t) = t^N.$$

Таким образом, кумулятивная функция распределения случайной величины Y равна

$$\mathrm{F}(t) = \begin{cases} 0, & \text{если } t \le 0 \\ t^N, & \text{если } 0 < t < 1 \\ 1, & \text{если } t \ge 1 \end{cases}$$

а плотность распределения случайной величины Y на интервале $(0,1)$ равна $\varphi(t) = F'(t) = Nt^{N-1}$. **(3 point)**

3) Теперь легко вычислить математическое ожидание случайной величины Y:

$$\mathrm{E}(Y) = \int_0^1 t \cdot Nt^{N-1} dt = \frac{N}{N+1}$$ **(0,5 point)**

Задание 11 (7 баллов).

<u>Математическая статистика №1</u>

Random variables $X_1, \dots, X_n$ are independent and identically distributed with the following probability density function:

$$f(x; a) = \begin{cases} a + 2x(1 - a), & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Here $a \in [0; 2]$ is an unknown parameter.

a) **(3,5 points)** Random variable $\hat{a} = m - 3(X_1 + X_2)$ is an unbiased estimator for a. What is the value of m?

b) **(4 points)** Derive the method of moments estimator for a.

Solution

(a) $\hat{a}$ is unbiased, so $E(\hat{a}) = m - 3\big(E(X_1) + E(X_2)\big) = a$.

We start with $E(X_i)$:

$$E(X_i) = \int_{-\infty}^{\infty} xf(x;a)dx = \int_0^1 (ax + 2x^2 - 2ax^2)dx = \left(\frac{ax^2}{2} + \frac{2x^3}{3} - \frac{2ax^3}{3}\right)\Bigg|_0^1$$

$$= \left(\frac{a}{2} + \frac{2}{3} - \frac{2a}{3}\right) - 0 = \frac{4-a}{6}.$$

And now the estimator:

$$E(\hat{a}) = m - 3 \cdot 2 \cdot \frac{4-a}{6} = m - 4 + a$$

Hence, $m = 4$.

(b) The moment equation $E(X_i) = \bar{X}$ turns into:

$$\frac{4-a}{6} = \bar{X}.$$

The solution is

$$\hat{a}_{MM} = 4 - 6\bar{X}.$$

Задание 12 (7 баллов).
Математическая статистика №2

A statistician estimates the mean of a normal population with known variance σ^2 from a sample that consists of n independent observations. Find the confidence level for the following intervals:

a) **(3,5 points)** $\bar{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2\frac{\sigma}{\sqrt{n}}$;

b) **(4 points)** $\bar{X} - 1.5\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2.5\frac{\sigma}{\sqrt{n}}$.

Here $\bar{X}$ denotes the sample mean.

Table. Values of standard normal cumulative distribution function

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224

0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Solution

The distribution of the sample mean: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Standardized sample mean: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

(a) $P\left(\bar{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2\frac{\sigma}{\sqrt{n}}\right) = P(-2 < Z < 2) = 0.9545.$

(b) $P\left(\bar{X} - 1.5\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2.5\frac{\sigma}{\sqrt{n}}\right) = P(-1.5 < Z < 2.5) = 0.927.$

Трек «Математические методы в прикладной социальной психологии»

Задание 13 (28 баллов).
1. Обоснованность, аргументированность, доказательность,
2. Конкретность и полнота ответа,
3. Умение анализировать данные на английском языке, извлекая суть проблемы,
4. Умение применить социально-психологические теоретические знания для анализа социальных ситуаций и поведения людей в них,
5. Упоминание фамилий специалистов при анализе и ссылках на теории,
6. Умение свободно отвечать на вопросы,
7. Отсутствие фактических ошибок.

За данное задание можно получить до 28 баллов.

Задание 14 (28 баллов).
1. Наличие в ответе авторской позиции по рассматриваемой тематике,
2. Обоснованность, аргументированность, доказательность высказываемых положений и выводов автора,
3. Знание социально-психологической проблематики и терминологии,
4. Умение применять социально-психологические теории и инструменты к анализу реальных явлений,
5. Видение прикладных аспектов социально-психологических теорий,
6. Отсутствие в ответе элементов обыденного психологического знания и журнализмов,
7. Упоминание фамилий специалистов при анализе и ссылках на теории,
8. Отсутствие фактических ошибок.

За данное задание можно получить до 28 баллов.
Максимальный балл – 56.

Трек «Математические методы в социологии»

Задание 15 (28 баллов).
Проведено исследование среди малых предпринимателей Москвы и Архангельска. На основе указанных респондентами **мотивов** создания своего дела были выделены три группы малых предпринимателей: **Профессионалы** (рассматривают дело как реализацию себя в своей профессии), **Инноваторы** (готовы к риску и хотят реализовать свои идеи) и **Семейный бизнес** (занимаются семейным делом, в частности, продолжают дело родителей).

В таблице даны основные параметры групп, а также средние оценки:
1. успешности их бизнеса:
«Как Вы оцениваете финансовый результат Вашего бизнеса в настоящее время?» (ответ по 5-балльной шкале, где 1 – очень сложное положение, … 5 – можно назвать очень успешным)
2. решения продолжать свой бизнес:
«Думали ли Вы о закрытии своего предприятия?» (ответ по 5-балльной шкале, где 1 – Никогда не было такой мысли, а 5 - Я уже почти принял(-а) это решение).

	Группы предпринимателей			**По всей выборке**
	Профессионалы	**Инноваторы**	**Семейный бизнес**	
N	54	47	41	142
Возраст (лет), среднее*	32,8	36,0	38,3	35,5

Стандартное отклонение	9,4	10,0	9,1	9,7
Стаж ведения бизнеса (лет), среднее*	5,1	7,1	8,6	6,7
Стандартное отклонение	3,9	6,6	8,02	6,3
Наличие образования, связанного со сферой предприятия (% по столбцу)	55,6	44,7	39,0	47,2
Оценка финансовой успешности своего бизнеса*				
Средняя оценок	4,02	3,38	3,61	3,60
Стандартное отклонение	0,85	0,94	0,99	0,95
Решение продолжать свое дело*				
Средняя оценок	1,8	2,02	2,34	2,00
Стандартное отклонение	1,2	1,18	1,17	1,19

* Переменные имеют нормальное распределение

Задание:

1) На основе имеющихся данных проверьте гипотезу о связи мотивов открытия своего бизнеса с его финансовой успешностью и решением его продолжать. При выборе критерия проверки гипотезы дайте подробное обоснование.
2) По итогам проверки гипотезы предложите свою интерпретацию полученных результатов.

КРИТЕРИИ ОЦЕНИВАНИЯ

точность, полнота, корректность ответа
Проверка гипотез предполагает расчеты о наличии значимой связи между переменными: 1) мотивы (это группы предпринимателей) и 2) Оценка финансовой успешности и решение продолжать свое дело.
В качестве критериев проверки гипотезы при имеющихся данных можно было выбрать ANOVA.
Пример расчета для связи мотивов и оценкой финансовой успешности (вторая подгипотеза рассчитывается аналогично):

F= MSbg/MSwg

MSbg=6,9/2=3,5
MSbg=168,2/139=1,21

F=2,9 < 3,1 (крит.) гипотеза о наличии значимой связи не подтверждается

аргументированность и обоснованность выводов и предложений
Аргументы выбора метода анализа и критериев проверки гипотезы должны обращаться к имеющимся данным. При описании результата проверки гипотез необходимо содержательное описание выводов, а не просто «гипотеза не подтверждается». Такая содержательная интерпретация также предполагает объяснения полученных результатов.

ясность, структурированность и логичность изложения решения и анализа
Оценивается последовательность и ясность расчетов, полнота аргументов и ответов на оба вопроса.

Задание 16 (28 баллов).
В рамках исследовательского проекта, посвященного влиянию Антитабачного законодательства на снижение распространения курения в России, социологи обратились к анализу **социально-экономических факторов** потребления табачных изделий. Для достижения поставленной цели была применена бинарная логистическая регрессия. В качестве **зависимой переменной** выступил следующий вопрос «Вы курите в настоящее время?» (ответ по дихотомической шкале, где 0 – «Нет, не курю», 1 – «Да, курю». Ряд характеристик (пол, наличие высшего образования, семейное положение, наличие проблем со здоровьем) также измерялись в дихотомической шкале (см. таблицу ниже), в то время как показатели «самооценка здоровья» (1 – «Хорошее здоровье», 2 – «Среднее здоровье» и 3 – «Плохое здоровье») и «удовлетворенность жизнью» (1 – «Удовлетворены жизнью», 2 – «Средне удовлетворены», 3 – «Не удовлетворены») – в порядковой.

Независимые переменные	**B**	**Стандартная ошибка**	**Вальд**	**Значимость**	**Exp(B)**
Мужской пол (женский пол = 0)	0,978	0,259	4,272	0,000	2,658
Возраст (в годах)	0,105	0,03	11,978	0,001	1,900
Наличие высшего образования (0 – отсутствие ВО)	-0,488	0,202	5,845	0,016	0,614
Состоит в браке (не состоит в браке = 0)	-0,388	0,23	2,852	0,091	0,678
Количество своих детей младше 18 лет	-0,057	0,007	8,607	0,000	0,745
Логарифм собственных доходов	0,408	0,098	1,227	0,000	1,504
Наличие проблем со здоровьем в последнее время (0 – нет проблем)	-0,259	0,16	4,071	0,006	0,326
Среднее здоровье	0,894	0,976	2,201	0,074	1,019
Плохое здоровье	-0,861	3,001	9,135	0,001	0,951
Средне удовлетворены жизнью	0,174	8,423	5,254	0,295	2,233
Не удовлетворены жизнью	0,963	0,942	3,984	0,032	1,728
Константа	-0,489	0,076	1,693	0,043	0,513

Задание:

1) Представьте общую спецификацию (уравнение) регрессионной модели. Уровень доверительной вероятности равен 0,95. Опишите контрольную группу. Что означает константа?
2) Дайте содержательную интерпретацию результатов регрессионной модели. Какие факторы повышают/снижают вероятность курения?
3) Представьте, что Респондент А имеет следующие характеристики: мужчина, 28 лет, не женат и нет детей, имеет высшее образование и ежемесячный доход в размере 100.000 рублей (lg=5), имеет хорошее здоровье и не имел серьезных проблем со здоровьем в

последнее время, средне удовлетворен жизнью. Рассчитайте для респондента с указанными характеристиками вероятность курения.
4) Какие факторы исследователи не учли в модели? Какие переменные Вы бы добавили в качестве предикторов? Предложите методику их измерения, сформулировав вопрос и варианты ответа к нему.

КРИТЕРИИ ОЦЕНИВАНИЯ

Всего за вопрос – 28 баллов
Из них:

За задание 1 – 10 баллов
От участника ожидается запись модели бинарной логистической регрессии в линейной, либо в экспоненциальной форме (с указанием выбора формы).
На основе представленных предикторов предполагается также описание социально-демографических характеристик контрольной группы. Контрольная группа – незамужние женщины, без высшего образования, не имеющие проблем со здоровьем (хорошее здоровье) и обладающие высоким уровнем удовлетворенности жизнью.
Наконец, от участника ожидается интерпретация константы. Необходимо указать выше ли шансы наступления события (потребления табачных изделий) или не наступления события для контрольной группы. Т.к. $b0<0$, следовательно, шансы наступления события меньше 1 и вероятность наступления события меньше вероятности не наступления.
За игнорирование ответов на отдельные вопросы задания, баллы будут снижаться.

За задание 2 – 6 баллов
Ожидается, что участник не только формально выделит статистически значимые предикторы, но и сможет их содержательно проинтерпретировать. Кроме того, предложенная регрессионная модель позволяет выделить наиболее сильные предикторы, влияющие на потребление табачных изделий. К таковым можно отнести пол, возраст, удовлетворенность жизнью и уровень собственных доходов. Фактически представленные факторы можно условно разделить на 2 группы: повышающие вероятность потребления табачных изделий (принадлежность к мужскому полу, возраст, собственные доходы, низкая удовлетворенность жизнью) и снижающие ее (наличие высшего образования, количество детей, наличие проблем со здоровьем, плохое здоровье). Ряд предикторов значим на уровне доверительной вероятности 90% (например, семейное положение и средний уровень здоровья). При желании их тоже можно обозначить.
В целом мы ожидаем, что участник увидит специфику модели и адекватно опишет полученные результаты.
Если участник демонстрирует умение читать результаты регрессионной модели и корректно его применяет, то он получает максимальный балл за задание (в зависимости от полноты и точности ответа). В случае небольших неточностей (например, игнорирование ряда переменных) в ответе, оценка **не будет превышать 4 баллов.**

За задание 3 – 6 баллов

От участника ожидается представить расчеты вероятности курения для индивида с перечисленными характеристиками. Предполагается использование следующей формулы:
$$p = \frac{1}{1+e^{z}}$$

За задание 4 – 6 баллов
От участника ожидается понимание специфики построения регрессионных моделей и связанных с ними методологических трудностей (например, проблемы эндогенности).

Предполагается, что участник не только предложит потенциальные факторы потребления табачных изделий, но и представит аргументацию их выбора. Не менее важным оказывается и возможная операционализация предложенного концепта (формулировка вопроса и вариантов ответа).
Если никакой аргументации по дополнению моделей не поступило, участник **получает максимум 4 балла за задание.**

Общий критерий оценки работы: работа должна быть оформлена в виде эссе (текста). Это предполагает литературное оформление, т.е. ясный язык, связанность предложений и абзацев, общую логическую линию повествования. За работу в виде отрывочных фраз, перечеркнанный, нечитаемый текст, будет снижаться оценка (вплоть **до 10 баллов**). Совсем нечитаемая работа получает **0 баллов**.

Задание 17 (28 баллов).
Основным исследовательским вопросом группы ученых было понять, **какие ключевые факторы влияют на удовлетворенность жизнью россиян**. Ниже приведена таблица с результатами линейной регрессии по итогам данного исследования.

Задание:

1. Оцените общее качество регрессионной модели.
2. Дайте содержательную интерпретацию связи между ключевыми независимыми и контрольными переменными и зависимой переменной.
3. Опишите, какие из выбранных параметров имеют наибольшее и наименьшее влияние на удовлетворенность жизнью россиян? Почему?
4. Проанализируйте релевантность используемого набора независимых переменных для решения исследовательской задачи. Предложите свой вариант усовершенствования имеющейся модели.

	Удовлетворенность жизнью *1 - абсолютно не удовлетворен(а), 7 - абсолютно удовлетворен(а)*						
	B	Стд. Ошибка	Станд. Бета коэф.	t	Знч.	Tolerance	VIF
Социально - демографические характеристики (контрольные переменные)							
Возраст (Количество лет)	-0,002	0,003	-0,013	-0,537	0,591	0,578	1,731
Пол (Мужской = 0, Женский = 1)	0,046	0,092	0,01	0,492	0,623	916	1,092
Образование (нет высшего образования = 0, есть высшее образование = 1)	0,101	0,066	0,048	1,522	0,128	0,357	2,799
Количество лет образования	-0,022	0,03	-0,024	-0,74	0,460	0,331	3,019
Общий доход домохозяйства в месяц (руб.)	-0,045	0,019	-0,043	-1,923	0,059	0,59	1,694
Удовлетворенность разными аспектами жизни							
Уровень счастья (1 - абсолютно несчастлив, 7 - абсолютно счастлив)	0,449	0,023	0,422	19,419	0,000	0,741	1,35
Удовлетворенность отношениями в семье (1 - абсолютно не удовлетворен(а), 7 - абсолютно удовлетворен(а))	0,086	0,027	0,064	3,225	0,001	0,88	1,136

Субъективная удовлетворенность своим здоровьем (1 - абсолютно не удовлетворен(а), 7 - абсолютно удовлетворен(а))	-0,19	0,071	-0,063	-2,692	0,007	0,645	1,55
Удовлетворенность своей работой (1 - абсолютно не удовлетворен(а), 7 - абсолютно удовлетворен(а))	0,154	0,028	0,137	5,543	0,000	0,568	1,76
Удовлетворенность своим материальным положением (1 - абсолютно не удовлетворен(а), 7 - абсолютно удовлетворен(а))	-0,507	0,065	-0,179	-7,816	0,000	0,665	1,505
$R^2 = 0,509$							

КРИТЕРИИ ОЦЕНИВАНИЯ

1. Дана оценка качества регрессионной модели, представлена содержательная интерпретация связи между ключевыми независимыми и контрольными переменными и зависимой переменной, выделены ключевые предикторы. Отрефлексирована представленная модель, оценена релевантность используемого набора независимых переменных для решения исследовательской задачи. Предложен обоснованный вариант усовершенствования имеющейся модели. **20-28 баллов** (в зависимости от допущенных ошибок и неточностей).

2. В отличие от п.1, нехватает содержательной интерпретации связи между ключевыми независимыми и контрольными переменными и зависимой переменной, указаны не все ключевые предикторы. Суждениям о релевантности используемого набора независимых переменных для решения исследовательской задачи не хватает аргументации. Предложения по усовершенствованию модели нерелевантны или им не хватает аргументации. **11-19 баллов** (в зависимости от допущенных ошибок и неточностей).

3. Дан релевантный ответ на 2 и менее заданий из 4-х или ни на одно из 4-х заданий не дано полного и аргументированного ответа. Ответ содержит много неточностей и неаргументированных суждений. **10 баллов и менее** (в зависимости от допущенных ошибок и неточностей)

Задание 18 (28 баллов).
Группа социологов решила изучить факторы самоорганизации собственников жилья в крупных городах. В качестве переменной, измеряющей эту самоорганизацию, была выбрана активность жильцов многоквартирных домов при решении общедомовых вопросов (доля жильцов, которые приходят на собрания и голосуют на них).
Для обследования были выбраны по 5 многоквартирных домов в 10 крупных городах России (элитные дома не обследовались). Сбор информации осуществлялся в 2020 году, но часть данных удалось собрать за 2015-2020 годы, в том числе данные о количестве голосовавших на собраниях, которые предоставили управляющие организации и инициаторы собраний. Данные за 5 лет агрегировали, т.к. число собраний за год не велико.
В результате, был построен ряд линейных регрессионных моделей (см. таблицу).
Переменную по забастовкам и акциям протеста сформировали по информации из городских СМИ и общению с городскими администрациями.

Данные по числу жильцов, «возрасту дома» и стоимости ЖКХ за 2015-2020гг. были предоставлены управляющими организациями.
Данные по числу жильцов с высшим образованием получили в результате сплошного опроса жильцов каждого из 50 обследованных домов, проведенного в 2020г.
По итогам сплошного опроса жильцов была смоделирована сетевая структура сообщества жильцов каждого дома и измерена степень центральности сети жильцов каждого из 50 обследованных домов.

Стандартизованные регрессионные коэффициенты
(зависимая переменная – средняя доля жильцов, участвовавших в голосованиях общедомовых собраний за 2015-2020 годы)

Независимые переменные	**модели**			
	1	**2**	**3**	**4**
Степень центральности сообщества жильцов дома, 2020г.	–	–	0,490** (3,647)	0,482** (3,104)
Среднегодовое число жильцов, имеющих право участвовать в голосовании, 2015-2020гг.	-0,644** (-3,171)	-0,778* (-3,642)	-0,458* (-2,554)	-0,498* (-2,402)
Доля жильцов с высшим образованием, имеющих право участвовать в голосовании, 2020г.	0,401* (1,988)	0,420* (2,068)	0,258 (1,470)	0,211 (1,115)
«Возраст дома» (от года сдачи дома в эксплуатацию), лет	0,580** (3,498)	0,434** (2,463)	0,360* (2,362)	0,298 (1,861)
Средняя плата за содержание и ремонт жилого помещения и коммунальные услуги на 1 квартиру дома, 2015-2020гг.	–	-0,334 (-2,024)	–	-0,179 (-1,169)
Среднегодовое число забастовок и акций протеста в городе, 2015-2020гг.	–	-0,136 (-0,950)	–	-0,003 (-0,023)
Число обследованных домов	50	50	50	50
R^2	0,343	0,373	0,529	0,521

Примечание. В скобках приведены значения t-статистики (двусторонний тест).
*p<0.05; **p<0.01

Задание:
1. Проинтерпретируйте полученные результаты. Какие содержательные выводы о самоорганизации сообществ можно из них сделать?
2. Если бы у Вас была возможность повторить данное исследование, то что бы Вы изменили, добавили и/или убрали из моделей? Почему?

3. В условии задачи не сказано, как измерялась и рассчитывалась переменная «степень центральности». На основе каких данных Вы бы ее измеряли? И как бы рассчитывали ее значение?

КРИТЕРИИ ОЦЕНИВАНИЯ

Всего за вопрос – 28 баллов
Из них:

За задание 1 – 10 баллов
Ожидается, что участник не только формально выделит статистически значимые предикторы, но и сможет их содержательно проинтерпретировать. Фактически представленные модели тестируют эффект от включения переменной «степень центральности» (модели 3 и 4 по сравнению с моделями 1 и 2). Эта переменная оказывается одним из самых сильных предикторов и уменьшает эффект от других значимых предикторов в моделях.
Содержательно, централизованная сеть способствует самоорганизации и коллективному действию. В данном случае, она облегчает сбор жильцов на собрания за счет лидера (группы лидеров).
Также мы ждем рассуждений о других значимых переменных: размеру и возрасту дома. Содержательно получается, что жильцы небольшого и старого дома более активны (но не во всех моделях). Связано это может быть с тем, что маленькое сообщество, которое длительное время проживает совместно, имеет больше шансов сформировать социальный капитал, способствующий коллективному действию.
Возраст дома в наибольшей степени «конфликтует» с переменной «степень центральности» - мы ожидаем, что участник это увидит и отрефлексирует в тексте. Например, предположит, что эти переменные связаны друг с другом, что ставит вопрос об их одновременном включении в модель. Это же касается и переменной «доля жильцов с высшим образованием».
В целом, мы ожидаем, что участник увидит описанные выше особенности моделей, адекватно их опишет и, по крайней мере, задаст вопросы, на которые представленная таблица ответов не дает (как бонус – выдвинет предположения, почему так происходит).
Простого словесного формального описания данных таблицы недостаточно. Необходимо содержательное обсуждение результатов, а не пересказ регрессионных коэффициентов. Если есть только формальное описание, то оценка **не будет превышать 2 баллов**.

За задание 2 – 8 баллов
Помимо того, что участник может предложить убрать незначимые предикторы, мы ожидаем рассуждений о возможной взаимосвязи переменных в представленных моделях. Также мы ждем аргументированных и обоснованных предложений по дополнению моделей. Недостаточно просто предложить измерить еще какую-либо переменную и включить в модель – необходимо пояснить, какой эффект от нее ожидается и почему.
Будет оцениваться способность участника выдвигать конструктивные предложения, а также способность их аргументировать.
Если никаких предложений по дополнению моделей не поступило, участник **получает максимум 3 балла** за задание. За слабую аргументацию предложений баллы будут снижаться.

За задание 3 – 10 баллов
От участника ожидается понимание сути степени центральности (centrality) точки в сети и сети в целом. Формулы помнить необязательно, но от участника ожидается понимание подходов к измерению центральности. Например, на основе степени точки (degree), ее

«срединности» (betweenness) или близости (closeness). Если участник демонстрирует такое понимание, то он получает **до 5 баллов** (в зависимости от полноты и точности ответа).
Далее от участника ожидается предложение (хотя бы один вариант) по тому, какую информацию необходимо собрать, чтобы измерить степень центральности. Предложение должно быть аргументировано. За это участник получает **до 3 баллов** (в зависимости от полноты и точности ответа).
Наконец, от участника ожидаются предложения по рассчету степени центральности сети. Мы ожидаем хотя бы общие принципы, подходы (формулы не требуются). Это оценивается **до 2 баллов**.

Общий критерий оценки работы: работа должна быть оформлена в виде эссе (текста). Это предполагает литературное оформление, т.е. ясный язык, связанность предложений и абзацев, общую логическую линию повествования. За работу в виде отрывочных фраз, перечерканный, нечитаемый текст, будет снижаться оценка (вплоть **до 10 баллов**). Совсем нечитаемая работа получает **0 баллов**.

Трек «Прикладная математика в инженерии и естественных науках»

Задание 19 (12 баллов).

Пусть A – квадртная матрица порядка n, элементы которой заданы условиями $a_{ij} = i^2 + j^2$.

a) Вычислить определитель матрицы A;

b) Вычислить $f'(0)$, если $f(x) = \det(xA + I)$, где I – единичная матрица порядка $n \in \mathbb{N}$, $x \in \mathbb{R}$.

Решение:

a) Нужно отдельно рассмотреть случаи $n = 1, 2$ и $n \geq 3$.

- Рассмотрим случай $n = 1$. Тогда $A = \begin{pmatrix} 2 \end{pmatrix}$ и $\det A = 2$.
- Рассмотрим случай $n = 2$. Тогда $A = \begin{pmatrix} 2 & 5 \\ 5 & 8 \end{pmatrix}$ и $\det A = 16 - 25 = -9$
- Рассмотрим случай $n \geq 3$. Тогда

$$A = \begin{pmatrix} 2 & 5 & 10 & \dots & 1+n^2 \\ 5 & 8 & 13 & \dots & 4+n^2 \\ 10 & 13 & 18 & \dots & 9+n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n^2+1 & n^2+4 & n^2+9 & \dots & 2n^2 \end{pmatrix}$$

При вычислении $\det A$ сначала из 3-ей строки вычтем 2-ую строку, а затем из 2-ой строки вычтем 1-ую.

Получаем

$$\det A = \begin{vmatrix} 2 & 5 & 10 & \dots & 1+n^2 \\ 3 & 3 & 3 & \dots & 3 \\ 5 & 5 & 5 & \dots & 5 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n^2+1 & n^2+4 & n^2+9 & \dots & 2n^2 \end{vmatrix}$$

$$= 3 \cdot 5 \cdot \begin{vmatrix} 2 & 5 & 10 & \dots & 1+n^2 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n^2+1 & n^2+4 & n^2+9 & \dots & 2n^2 \end{vmatrix} = 0\,,$$

так как в определителе 2-я и 3-я строки совпадают.

b) Заметим, что $f(x) = \sum\limits_{k=0}^{n} b_k x^k$. Тогда $f'(0) = b_1$.

Легко видеть, что $b_1 = \operatorname{tr} A = \sum\limits_{i=1}^{n} a_{ii}$.

Остается заметить, что $a_{ii} = i^2 + i^2 = 2i^2$.
И окончательно получаем

$$f'(0) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} 2i^2 = 2 \cdot \frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)(2n+1)}{3}$$

Критерии оценивания:

Пункт a)

6 баллов : Полностью обоснованное решение с правильным ответом;

4-5 баллов : Правильный решение с несущественными погрешностями в его обосновании или незначительными опечатками;

1-3 баллов : Правильно решена только часть задачи. Например, рассмотрен только случай $n \geq 3$ или только случаи $n = 1$ и/или $n = 2$.

Пункт b)

6 баллов : Полностью обоснованное решение с правильным ответом;

4-5 баллов : Правильный решение с несущественными погрешностями в его обосновании или незначительными опечатками. Например, не вычислена сумма $\sum\limits_{i=1}^{n} i^2$;

1-3 баллов : Различные попытки решения. Высказаны идеи, ведущие к решению.

Задание 20 (14 баллов).

Найдите решение $y = y(x)$ дифференциального уравнения:

$$y'' + 2y' + 2y = e^{-x}$$

при условии $y(0) = 1$ и $\int_0^{+\infty} y(x)\,dx = 0$.

Решение

Рассмотрим однородное дифференциальное уравнение:

$$y_o'' + 2y_o' + 2y_o = 0.$$

Это линейное уравнение с постоянными коэффициентами. Найдем корни характеристического многочлена:

$$\lambda^2 + 2\lambda + 2 = 0 \quad \Rightarrow \quad \lambda_{1,2} = -1 \pm i.$$

Следовательно, общее решение однородного уравнения имеет вид:

$$y_o = C_1 e^{-x} \sin(x) + C_2 e^{-x} \cos(x).$$

Правая часть неоднородного уравнения является квазимногочленом. Частное решение неоднородного уравнения можно искать в виде Ce^{-x}. Подставляя частное решение Ce^{-x} в исходное уравнение, получаем:

$$(C - 2C + 2C)e^{-x} = e^{-x} \quad \Rightarrow \quad C = 1.$$

Таким образом, общее решение исходного неоднородного уравнения имеет вид суммы общего решения однородного уравнения и частного решения неоднородного уравнения:

$$y = C_1 e^{-x} \sin(x) + C_2 e^{-x} \cos(x) + e^{-x}.$$

Остается найти константы $C_{1,2}$, подставляя решение в начальные условия. Условие в точке $x = 0$ дает:

$$y(0) = 1 \quad \Rightarrow \quad C_2 + 1 = 1 \quad \Rightarrow \quad C_2 = 0.$$

Вычисление интеграла из второго условия:

$$\int_0^{+\infty} y(x)\,dx = \int_0^{+\infty} (C_1 \sin(x) + 1)e^{-x}\,dx = \frac{C_1}{2} + 1 = 0 \quad \Rightarrow \quad C_1 = -2.$$

Решение имеет вид:

$$y = (1 - 2\sin(x))e^{-x}.$$

Критерии задачи

14 баллов: Полностью обоснованное решение с правильным ответом.
11-13 баллов: Правильный решение с несущественными погрешностями в его обосновании или незначительными опечатками.
8-10 баллов: Правильный подход к решению задачи, но допущены ошибки или пропущены существенные блоки в обосновании решения. Например, есть ошибки при взятии интеграла или построении частного решения.
5-7 баллов: Правильно решена только часть задачи. Например, построено общее решение дифференциального уравнения, но не найдены неопределенные константы.
1-4 баллов: Различные попытки решения. Высказаны идеи, ведущие к решению.

Задание 21 (14 баллов).

Ниже представлена программа на языке программирования Python. Что выведет данная программа?

In [7]:

```
def fun(num):
    lst = [1, 1]
    n = 0
    while len(lst) < num:
        lst.append(1)
        lst.append(1)
        n += 1
        for k in range(n):
            term = lst[-n-1] + lst[-n-2]
            lst.append(term)
    return lst[num-1]

print(fun(2021) / fun(2020))
```

Решение

Функция fun заполняет список lst элементами треугольника Паскаля, записанного по строкам. Результатом вызова функции является значение биномиального коэффициента, определяемого порядковым номером элемента в треугольнике Паскаля.

Таким образом, для нахождения результата вызова fun(2021) необходимо определить в какой стороне треугольника Паскаля и с каким порядковым номером будет находиться данный элемент.

Сначала найдем число полностью заполненных строк. В k-й строке треугольника Паскаля содержится k+1 элементов, то есть полностью m заполненных строк содержат

$$1 + 2 + 3 + \ldots (m + 1) = m(m + 1)/2$$

элементов. Легко получить, что 2020-й и 2021-й элементы будут находиться в 64-й строке (m = 63), т.к. $63 \times 64/2 = 2016$.

Нужные нам элементы расположены в 64-й строке на 4-м и 5-м местах. Таким образом, результатом работы программы будет отношение биноминальных коэффициентов

$$\frac{C_{63}^4}{C_{63}^3} = 15$$

Критерии оценивания

- 14 баллов: Полное решение с обоснованным ответом
- 13-11 баллов: Незначительные погрешности в обосновании ответа
- 8-10 баллов: в целом верное решение с допущенными ошибками
- 5-7 баллов: высказаны некоторые идеи, ведущие к правильному решению
- 1-4 баллов: различные попытки построить или угадать решение

Задание 22 (16 баллов).

На некотором пешеходном переходе установлен светофор для пешеходов, который горит либо зеленым светом (разрешено движение для пешеходов), либо красным (движение для пешеходов запрещено). Далее мы предполагаем, что время измеряется в минутах.

Известно, что если в момент времени **t** включен зеленый, то вероятность того, что зеленый переключится на красный на интервале времени (t, t+δ)(δ↓0), равна δ+o(δ) и не зависит от предыстории процесса.
Если в момент времени **t** включен красный, то вероятность того, что красный переключится на зеленый на интервале времени (t, t+δ)(δ↓0), равна $\frac{1}{2}$δ+o(δ) и не зависит от предыстории процесса.

1. Если сейчас светофор зеленый, то какова вероятность того, что через 2 минуты пешеход увидит зеленый свет?
2. Предположим, что 1 января 2021 года в 8 часов горел зеленый свет. Найти приближенную вероятность того, что 1 января 2022 года в 8 часов пешеход увидит красный свет светофора (при условии, что весь год светофор работает в том же режиме без поломок и сбоев).

Решение

Введём случайный процесс X(t), который принимает два значения 0 и 1: X(t) = 0, если в момент времени t горит красный свет, X(t) = 1, если в момент времени t горит зелёный свет. В силу условий задачи, процесс X(t) – однородный марковский с матрицей интенсивностей

$$Q = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{pmatrix}$$

Обозначим p_0 (t) – вероятность того, что X(t) = 0, т.е. в момент времени t горит красный свет, а через p_1 (t) обозначим вероятность того, что X(t) = 1, т.е. в момент времени t горит зелёный свет. p_0 (t) + p_1 (t) = 1 для всех t ≥ 0. Тогда в п.1. требуется найти p_1 (2). Запишем систему дифференциальных уравнений Колмогорова:

$$p_0'(t) = -\frac{1}{2}p_0(t) + p_1(t)$$

$$p_0'(t) = -\frac{1}{2}p_0(t) - p_1(t)$$

с начальными условиями $p_0(0) = 0$, $p_1(0) = 1$. Подставим во второе уравнение $p_0(t) = 1 - p_1(t)$. Получим

$$p_1'(t) = \frac{1}{2} - \frac{3}{2}p_1(t)$$

Это линейное неоднородное уравнение, общее решение которого имеет вид

$$p_1(t) = \frac{1}{3} + Ce^{-\frac{3}{2}t}$$

Используем начальные условия $p_0(0) = 1$, получим $C = \frac{2}{3}$. Таким образом,

$$p_1(t) = \frac{1}{3} + \frac{2}{3}e^{-\frac{3}{2}t}$$

п.1. Если сейчас светофор зелёный, то вероятность того, что через 2 минуты пешеход увидит зелёный свет, равна

$$p_1(2) = \frac{1}{3} + \frac{2}{3}e^{-3} \approx 0{,}367$$

п.2. Приближенная вероятность того, что 1 января 2022 года в 8 часов пешеход увидит красный свет светофора (при условии, что весь год светофор работает в том же режиме без поломок и сбоев), есть

$$\lim_{t\to\infty} p_0(t) = 1 - \lim_{t\to\infty} p_1(t) = \frac{2}{3} \approx 0{,}667$$

Ответ: п.1. $\frac{1}{3} + \frac{2}{3}e^{-3} \approx 0{,}367$; п.2. $\frac{2}{3} \approx 0{,}667$

Критерии:

16 баллов: полное обоснованное решение
15-14 баллов: в целом, верное решение. Небольшие погрешности
13-10 баллов: в целом, верное решение, но неполное и недостаточно обоснованное, или содержит ошибки
9-8 баллов: найдено решение задачи
7-5 баллов: высказаны некоторые идеи, ведущие к правильному решению
4-1 баллов: различные попытки построить или угадать решение
0 – нет решения, нет верных ответов, нет идей или попыток, которые могут привести к верному ответу