

Задача 1. Да се симулира хвърляне на зар и да се намери приближение на вероятността за падане на шестица.

Задача 2. В отдел на фирма работят 20 човека. За Коледа те решават да си разменят подаръци. В кутия слагат 20 листчета, на всяко от които има едно име. Всеки тегли листче (без да го връща) и подарява на този, чието име е изтеглил. Каква е вероятността поне един да изтегли своето име?

Задача 3. Каква е вероятността в група от 25 човека поне двама да имат рожден ден на един и същи ден от годината?

Задача 4. Студент се явява на изпит с конспект от 20 въпроса. От тях не знае само 3 въпроса. На изпита си тегли 2 въпроса от конспекта. Каква е вероятността да знае само един от изтеглените въпроси?

Задача 5. Иван има 5 ключа, но не знае кой е за неговата стая. Той пробва последователно с всеки от тях, като помни кой ключ е пробвал. Каква е вероятността да отключи с петия ключ?

Задача 6. На всеки от върховете на равностранен триъгълник има една мравка. Всяка мравка избира произволно един от другите два върха и тръгва към него. За единица време всяка мравка изминава разстоянието от един връх до друг. Две мравки могат да се разминат ако тръгнат една срещу друга. Каква е вероятността след единица време да има по една мравка на всеки връх?

Задача 7. Имаме 3 карти: първата е бяла от двете страни, втората е черна от двете страни, а третата е бяла от едната и черна от другата страна. Всяка карта е поставена в затворена кутия. Избираме произволна кутия, отваряме я и виждаме, че горната страна на картата в нея е бяла. Каква е вероятността другата страна на картата също да е бяла?

Задача 8. В кутия има 5 топки, номерирани от 1 до 5. Вадим произволна топка и я връщаме в кутията. Отново вадим произволна топка. Каква е вероятността да извадим два пъти една и съща топка?

Задача 9. В кутия има 3 различни чифта чорапи. Вадим в тъмното 2 чорапа. Каква е вероятността да са чифт?

Задача 10. Група от 20 човека, измежду които са Иван и Георги, е подредена по случаен начин в редица. Каква е вероятността Иван и Георги да са един до друг?

Задача 11. Тесте от 52 карти е разбъркано и е раздадено на 4 играчи. Каква е вероятността всеки играч да има едно асо?

Задача 12. Парола, състояща се от 5 малки букви, е генерирана по случаен начин (буквата на всяка позиция е избрана равновероятно измежду буквите a, b, \dots, z). Каква е вероятността всички букви в паролата да са различни?

Задача 13. На първия етаж на административна сграда 7 души чакат асансьора. Всеки от тях отива в някой от офисите в сградата. Сградата има 16 етажа и на всеки етаж има равен брой офиси (на първия етаж няма офиси).

а) Каква е вероятността поне двама от чакащите да отиват на един и същи етаж?

б) Ако Вие сте един от седемте, каква е вероятността поне един от останалите 6 да отива на Вашия етаж?

* * *

Задача 14. Хвърляме зар 10 пъти. Каква е вероятността да се паднат само 2 шестници? Каква е вероятността да се паднат не повече от 2 шестници? Каква е вероятността да се паднат 2 или повече шестници?

Задача 15. Хвърляме зар докато се падне шестика. Каква е вероятността да хвърляме не повече от 10 пъти? Каква е вероятността да хвърляме поне 6 пъти?

Задача 16. Хвърляме зар докато се паднат три шестници. Каква е вероятността да хвърляме не повече от 20 пъти?

Задача 17. Фенерче работи с 2 батерии. Иван има 8 батерии, от които 5 са нови и 3 са изтощени, но не знае кои точно. Ако пробва с 2 случайно избрани батерии, каква е вероятността фенерчето да проработи?

Задача 18. Машинописка прави средно по една грешка на всеки 500 думи. На една страница има 300 думи. Каква е вероятността машинописката да направи не повече от 2 грешки на 5 страници? Каква е вероятността да направи между 1 и 3 грешки (включително) на 5 страници?

Задача 19. На студенти е даден тест от 10 въпроса, всеки с по 4 възможни отговора, един от които е верен. Иван се явява на теста без да е учил и огражда произволно отговори. Каква е вероятността да е отговорил вярно на поне 5 от въпросите?

Задача 20. За клинично проучване са необходими доброволци имачи определен ген, който се среща с вероятност 0.1. Каква е вероятността да се тестват 5 или повече доброволци до намиране на първия доброволец с въпросния ген? Каква е вероятността да се тестват 50 или повече доброволци до намиране на 10 доброволци с въпросния ген?

Задача 21. Средно веднъж на 90 дни в софийското метро възниква технически проблем, който води до спиране на движението на влаковете за поне 20 минути. Каква е вероятността за 360 дни да възникне такъв проблем повече от 3 пъти?

Задача 22. Теглим 10 случайно избрани карти от тесте с 52 карти (без връщане). Каква е вероятността да изтеглим поне 2 купи?

Задача 23. Теглим 10 пъти по една случайно избрана карта от тесте с 52 карти (с връщане). Каква е вероятността да изтеглим поне 2 купи?

Задача 24. Пациент в болница се нуждае от кръводарител с кръвна група АВ. Известно е, че 7% от населението е с кръвна група АВ. Потенциални кръводарители се тестват в болницата за определяне на кръвната им група.

а) Каква е вероятността да се тестват не повече от 10 души до откриване на кръводарител с кръвна група АВ?

б) Каква е вероятността измежду първите 50 тествани да има поне двама с кръвна група АВ?

Задача 25. Броят клиенти, които посещават офис на банка за един ден, е Пуасоново разпределен с параметър $\lambda = 18$.

а) Каква е вероятността офиса на банката да бъде посетен от поне 20 клиенти за един ден?

б) Каква е вероятността между 15 и 25 клиенти (включително) да посетят офиса на банката за един ден?

* * *

Задача 26. Генерирайте 500 случайни числа от равномерно разпределение в интервала $(2, 3)$. Начертайте хистограма на генерираните числа и на същата картинка добавете графика на плътността $f(x)$. Повторете същото с 5000 случайни числа.

Задача 27. Генерирайте 500 случайни числа от експоненциално разпределение с параметър $\lambda = 1/7$. Начертайте хистограма на генерираните числа и на същата картинка добавете графика на плътността $f(x)$. Повторете същото с 5000 случайни числа.

Задача 28. Генерирайте 500 случайни числа от нормално разпределение с параметри $\mu = 0$, $\sigma = 1$. Начертайте хистограма на генерираните числа и на същата картинка добавете графика на плътността $f(x)$. Повторете същото с 5000 случайни числа.

Задача 29. Количеството (в милилитри) на душ гел в опаковка е равномерно разпределено в интервала $(248, 255)$. Каква е вероятността произволно избрана опаковка да съдържа по-малко от 250 мл душ гел? Намерете v , такова че произволно избрана опаковка съдържа поне v мл душ гел с вероятност 0.95.

Задача 30. Времето на живот (в хиляди часове) на лазерен диод е експоненциално разпределено с параметър $\lambda = 1/10$. Каква е вероятността времето на живот на произволно избран лазерен диод да надвишава 10 хиляди часа? Каква е вероятността времето на живот на произволно избран лазерен диод да е между 7 и 11 хиляди часа? Намерете t , такова че времето на живот на лазерен диод е поне t с вероятност 0.97.

Задача 31. Времето от зареждане до изтощаване на батерия на лаптоп при обичайна работа е нормално разпределено със средно 260 минути и стандартно отклонение 50 минути. Каква е вероятността батерия да се изтощи след повече от 4 часа работа? Каква е вероятността батерия да се изтощи след между 3 и 5 часа работа? Намерете t , такова че батерия се изтощава след повече от t минути с вероятност 0.9.

Задача 32. Количеството кашкавал, което се изразходва в дадена пицария за една седмица, е нормално разпределено със средно 41 кг и стандартно отклонение 5 кг. Каква е вероятността в произволно избрана седмица пицарията да изразходи над 51 кг кашкавал? Колко кашкавал трябва да има в запас в дадена седмица, така че да е достатъчно с вероятност 0.99?

Задача 33. Количеството бира (в литри), което Иван изпива за една седмица, е експоненциално разпределено с параметър $1/3$. Каква е вероятността Иван да изпие повече от 10 литра бира за една седмица? Каква е вероятността в най-много 3 от следващите 5 седмици Иван да изпие повече от 10 литра бира? Предполагаме, че количеството изпита бира в дадена седмица е независимо от това в предишните седмици.

Задача 34. Обемът течност, който автомат за безалкохолни напитки налива в една чаша, е нормално разпределен със средно 205 мл и стандартно отклонение 2 мл. Каква е вероятността автоматът да налее в една чаша по-малко от 200 мл? Каква е вероятността автоматът да налее в една чаша между 200 и 205 мл?

Задача 35. Времето, през което потребител разглежда интернет страница преди да премине на друга страница, е експоненциално разпределено със средно 5 секунди. Каква е вероятността страница да бъде разглеждана от потребител повече от 10 секунди? След колко време 50% от потребителите са преминали на друга страница?

* * *

Задача 36. Разгледайте данните [survey](#) от пакета [MASS](#). Представете таблично и графично данните за физическите упражнения ([Exer](#)).

Задача 37. (Данни [survey](#) от пакета [MASS](#).) Представете таблично и графично данните за пушенето ([Smoke](#)).

Задача 38. (Данни [survey](#) от пакета [MASS](#).)

- а) Какъв процент от студентите са непушачи?
- б) Колко от студентите са непушачи и правят физически упражнения често?
- в) Какъв процент от студентите са непушачи и правят физически упражнения често?
- г) Какъв процент от непушачите правят физически упражнения често?
- д) Какъв процент от студентите, правещи физически упражнения често, са непушачи?

Задача 39. (Данни [survey](#) от пакета [MASS](#).) Чрез подходящи графики илюстрирайте честотата на пушене в зависимост от честотата на физически упражнения.

Задача 40. (Данни [survey](#) от пакета [MASS](#).) Чрез подходящи графики илюстрирайте честотата на физически упражнения в зависимост от честотата на пушене.

Задача 41. (Данни [survey](#) от пакета [MASS](#).) Представете чрез подходящи таблици и графики данните за пулса на студентите ([Pulse](#)).

Задача 42. (Данни `survey` от пакета `MASS`.) Представете чрез подходящи таблици и графики данните за възрастта на студентите (`Age`).

* * *

Задача 43. (Данни `survey` от пакета `MASS`.) Намерете медианата, средната стойност и стандартното отклонение на:

- а) пулса на студентите;
- б) пулса на жените;
- в) пулса на студентите на възраст не повече от 25 години;
- г) пулса на студентите, правещи физически упражнения често;
- д) пулса на студентите, които са непушачи и правят физически упражнения често.

Задача 44. (Данни `survey` от пакета `MASS`.) Чрез подходящи числови характеристики и графики покажете как пулсът се различава в зависимост от честотата на физически упражнения (`Exer`).

Задача 45. Във файла `exam_grades.txt` има данни за оценките на студенти от курс по статистика.

- а) Чрез графики и числови характеристики изследвайте зависимостта между:
 - `test1` и `course.grade`;
 - `test2` и `course.grade`;
 - `test3` и `course.grade`.
- б) Чрез графики и числови характеристики покажете как крайната оценка (`course.grade`) се различава в зависимост от пола.
- в) Покажете как крайната оценка варира в зависимост от семестъра.

Задача 46. Генерирайте вектор $\mathbf{x} = (x_1, \dots, x_{500})$ с 500 случайни числа от:

- а) $U(3, 7)$;
- б) $\text{Exp}(1/5)$;
- в) $\mathcal{N}(5, 1)$.

За всеки от векторите постройте кутия с мустаци, ориентирана хоризонтално. Добавете на графиката \bar{x} и границите на интервала $[\bar{x} - 3s, \bar{x} + 3s]$ като използвате функцията `points`.

Задача 47. (Данни `survey` от пакета `MASS`.) Намерете медианата, средната стойност и стандартното отклонение на:

- а) възрастта на студентите;
- б) възрастта на пушачите;
- в) възрастта на пишещите с дясната ръка;
- г) възрастта на студентите с измерен пулс 70 и повече;
- д) възрастта на студентите, които не правят физически упражнение.

Задача 48. (Данни `survey` от пакета `MASS`.) Представете графично височината на студентите. Чрез подходящи числови характеристики и графики сравнете височината на мъжете и жените.

* * *

Задача 49. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат нормално разпределение $\mathcal{N}(\mu, \sigma^2)$. Намерете оценка на вероятността $\theta_a = \mathbb{P}(X_i \leq a)$, като

- (1) използвате, че x_1, x_2, \dots, x_n са генерирани от нормално разпределение;
- (2) не използвате, че x_1, x_2, \dots, x_n са генерирани от нормално разпределение.

За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $\mu = 5$, $\sigma = 1$ и пресметнете двете оценки на вероятността θ_a за $a = 3.4$ и $a = 4.5$. Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

Задача 50. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат равномерно разпределение $U(0, b)$. Оценка на параметъра b може да се намери по метода на максималното правдоподобие и тази оценка е $\hat{b} = \max\{x_1, \dots, x_n\}$. Нека M е медианата на X_i , т.е. $\mathbb{P}(X_i \leq M) = 1/2$. Не е трудно да се види, че $M = b/2$. Ако използваме оценката $\hat{b} = \max\{x_1, \dots, x_n\}$, ще получим оценка на медианата $\hat{b}/2 = \max\{x_1, \dots, x_n\}/2$. Друга оценка на медианата M е извадъчната медиана.

За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $b = 8$ и пресметнете двете оценки на медианата. Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

Задача 51. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат експоненциално разпределение $\text{Exp}(\lambda)$. Оценка на параметъра λ може да се намери по метода на максималното правдоподобие и тази оценка е $\hat{\lambda} = 1/\bar{x}$. Нека M е медианата на X_i , т.е. $\mathbb{P}(X_i \leq M) = 1/2$. Не е трудно да се види, че $M = (1/\lambda) \log(2)$. Ако използваме оценката $\hat{\lambda} = 1/\bar{x}$, ще получим оценка на медианата $(1/\hat{\lambda}) \log(2) = \bar{x} \log(2)$. Друга оценка на медианата M е извадъчната медиана.

За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $\lambda = 1/5$ и пресметнете двете оценки на медианата. Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

Задача 52. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат нормално разпределение $\mathcal{N}(\mu, \sigma^2)$. Разглеждаме следните оценки на σ :

$$\hat{\sigma}_1 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\sigma}_2 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

а) За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $\mu = 3$, $\sigma = 2$ и пресметнете двете оценки на σ . Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

б) Може да разглеждаме $\hat{\sigma}_1^2$ и $\hat{\sigma}_2^2$ като оценки на σ^2 . За всяко n пресметнете средното и стандартното отклонение на двете оценки на σ^2 на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

Задача 53. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат експоненциално разпределение $\text{Exp}(\lambda)$. Оценка на дисперсията $\text{Var}(X_i) = 1/\lambda^2$ може да се намери като се използва оценката $\hat{\lambda} = 1/\bar{x}$. Друга оценка на дисперсията $\text{Var}(X_i)$ е s^2 .

За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $\lambda = 1/5$ и пресметнете двете оценки на дисперсията. Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

Задача 54. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат експоненциално разпределение $\text{Exp}(\lambda)$. Намерете оценка на вероятността $\theta_a = P(X_i \leq a)$, като

- (1) използвате, че x_1, x_2, \dots, x_n са генерирани от експоненциално разпределение;
- (2) не използвате, че x_1, x_2, \dots, x_n са генерирани от експоненциално разпределение.

За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $\lambda = 1/5$ и пресметнете двете оценки на вероятността θ_a за $a = 0.3$ и $a = 1.8$. Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

Задача 55. Нека x_1, x_2, \dots, x_n са наблюдения над случайните величини X_1, X_2, \dots, X_n , които са независими и имат Поасоново разпределение $\text{Po}(\lambda)$. Намерете оценка на дисперсията $\text{Var}(X_i)$, като

- (1) използвате, че x_1, x_2, \dots, x_n са генерирани от Поасоново разпределение;
- (2) не използвате, че x_1, x_2, \dots, x_n са генерирани от Поасоново разпределение.

За $n = 20, 50, 100, 500, 1000$ генерирайте данни x_1, x_2, \dots, x_n като вземете $\lambda = 3$ и пресметнете двете оценки на дисперсията. Повторете 30000 пъти. За всяко n пресметнете средното и стандартното отклонение на двете оценки на базата на получените 30000 стойности. Визуализирайте получените резултати на подходяща графика.

* * *

Задача 56. За $n = 3, 7, 10, 30, 90, 200$ генерирайте данни x_1, x_2, \dots, x_n от експоненциално разпределение с параметър $\lambda = 1/5$ и пресметнете $\Sigma = x_1 + \dots + x_n$. Повторете 10000 пъти.

$$\begin{array}{cccc} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(10000)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(10000)} \\ \vdots & \vdots & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(10000)} \\ \Sigma^{(1)} & \Sigma^{(2)} & \dots & \Sigma^{(10000)} \end{array}$$

За всяко n начертайте хистограма на $\Sigma^{(1)}, \Sigma^{(2)}, \dots, \Sigma^{(10000)}$.

Задача 57. За $n = 3, 7, 10, 30, 90, 200$ генерирайте данни x_1, x_2, \dots, x_n от експоненциално разпределение с параметър $\lambda = 1/5$ и пресметнете $\bar{x} = (x_1 + \dots + x_n)/n$. Повторете 10000 пъти.

$$\begin{array}{cccc} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(10000)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(10000)} \\ \vdots & \vdots & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(10000)} \\ \bar{x}^{(1)} & \bar{x}^{(2)} & \dots & \bar{x}^{(10000)} \end{array}$$

За всяко n начертайте хистограма на $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(10000)}$.

Задача 58. Решете предходната задача като генерирайте данни от равномерно разпределение в интервала $(2, 8)$.

Задача 59. Измерена е дължината на дясното ухо на 66 случайно избрани жени на възраст между 18 и 30 години. Резултатите показват средна дължина 61.9 мм и стандартно отклонение 4.1 мм. Нека предположим, че популационната дисперсия е известна и $\sigma^2 = (4.1)^2$.

а) Намерете 95-процентен доверителен интервал за средната дължина на дясното ухо на жените на възраст между 18 и 30 години.

б) Ако са направени измервания на 88 жени и са получени същите резултати за средната дължина и стандартното отклонение, намерете 95-процентен доверителен интервал за средната дължина на дясното ухо на жените на възраст между 18 и 30 години.

Задача 60. Кадмийт е тежък метал, който може да се натрупа до високи нива в гъбите. Измерена е концентрацията на кадмий в случайна извадка от 10 горски гъби събрани в дадена местност. Резултатите са следните (в мг/кг):

3.1 3.0 3.7 2.6 4.2 3.8 3.6 2.7 3.8 4.4

Може да предположим, че концентрацията на кадмий е нормално разпределена.

а) Намерете 95-процентен доверителен интервал за средната концентрация на кадмий в горските гъбите в дадената местност.

б) Намерете 90-процентен доверителен интервал за средната концентрация на кадмий в горските гъбите в дадената местност.

Задача 61. Проведен е експеримент, при който 38 домашни гълъба са пуснати на свобода от непознато място, което е на разстояние 106 км от дома им. От тях 22 намерили пътя към дома си. Намерете 95-процентен доверителен интервал за вероятността гълъб от дадения вид да намери пътя към дома си от непознато място на 106 км.

Задача 62. Направено е проучване с цел да се изследва ефектът на звука от сърдечния ритъм на майката върху новороденото. Бебетата в родилно отделение са разделени на две групи. Първата група е непрекъснато изложена на звука от сърдечния ритъм на възрастен, а втората група не е изложена на такъв звук. Измерена е промяната в теглото на бебетата от раждането до четвъртия ден. Данните са във файла [salk.txt](#). Намерете 95-процентен доверителен интервал за медианата на промяната в теглото на бебетата (за първата и за втората група).

Задача 63. С данните от задача 62 намерете 95-процентен доверителен интервал за разликата на медианите на първата и втората група.

* * *

Задача 64. Измерена е дължината на дясното ухо на 66 случайно избрани жени на възраст между 18 и 30 години. Резултатите показват средна дължина 61.9 мм и стандартно отклонение 4.1 мм. Нека предположим, че популационната дисперсия е известна и $\sigma^2 = (4.1)^2$. Имаме ли основание да твърдим, че средната дължина на дясното ухо на жените на възраст между 18 и 30 години е различна от 60 мм?

Задача 65. Кадмийт е тежък метал, който може да се натрупа до високи нива в гъбите. Измерена е концентрацията на кадмий в случайна извадка от 10 горски гъби събрани в дадена местност. Резултатите са следните (в мг/кг):

3.1 3.0 3.7 2.6 4.2 3.8 3.6 2.7 3.8 4.4

Може да предположим, че концентрацията на кадмий е нормално разпределена. Имаме ли основание да твърдим, че средната концентрация на кадмий в горските гъбите в дадената местност е по-малко от 4 мг/кг?

Задача 66. Проведен е експеримент, при който 38 домашни гълъба са пуснати на свобода от непознато място, което е на разстояние 106 км от дома им. От тях 22 намерили пътя към дома си. Преди провеждане на експеримента изследовател твърдял, че над 51 на сто от гълъбите ще намерят пътя към дома си. Подкрепят ли данните твърдението на изследователя?

Задача 67. Според проучване от 2009 година, около 7.5% от стоките в даден хипермаркет имат грешна цена на етикета. През 2010 е направено ново проучване и от 200 случайно избрани стоки, 14 били с грешна цена.

а) Имаме ли основание да твърдим, че процентът на стоки с грешна цена се е променил спрямо 2009 година?

б) Имаме ли основание да твърдим, че процентът на стоки с грешна цена е намалял спрямо 2009 година?

Задача 68. Според производител на автомат за безалкохолни напитки, средното количество, което налива автоматът в една чаша е 170 грама със стандартно отклонение 3.9 грама. За да се провери това, е измерено количеството течност при 50 наливания. Резултатите показват средно 168 грама и същото стандартно отклонение. Може ли да се твърди, че в средно автоматът налива по-малко от 170 грама?

Задача 69. Измерено е нивото на хемоглобин на 10 деца страдащи от специфична болест. Данните са:

12.3 11.4 14.2 15.3 14.8 13.8 11.1 15.1 15.8 13.2

Считаме, че нивото на хемоглобин има нормално разпределение.

а) Имаме ли основание да твърдим, че средното ниво на хемоглобин при децата страдащи от тази болест е различно от нормалното ниво от 14.6?

б) Имаме ли основание да твърдим, че средното ниво на хемоглобин при децата страдащи от тази болест е по-малко от 14.6?

Задача 70. Преподавател дал два варианта (А и В) задачи на изпит. Резултатите от изпита са във файла `examAB.txt`. Може ли да се твърди, че в средно вариант В е по-труден от вариант А, т.е. в средно студентите получават по-малко точки ако им се е падне вариант В?

Задача 71. Проведен е експеримент, при който е измерено времето на реакция при 32 доброволци преди и след изпиване на 50 мл водка. Данните са във файла `reacttime.txt`. Може ли въз основа на данните да се заключи, че след употреба на алкохол времето на реакция в средно се увеличава?

Задача 72. Направена е извадка от 200 болта, произведени от машина А и 200 болта, произведени от машина В. Дефектни се оказали 8 болта произведени от първата машина и 15 болта произведени от втората. Може ли да се твърди, че двете машини се различават по отношение на вероятността да произведат дефектен болт?

Задача 73. За определяне на съдържанието на примеси в сплави на стомана съществуват два метода. В 8 проби е определено съдържанието на примеси по всеки от двата метода. Данните са следните:

Проба	Метод 1	Метод 2
1	1.2	1.4
2	1.3	1.7
3	1.5	1.5
4	1.4	1.3
5	1.7	2.0
6	1.8	2.1
7	1.4	1.7
8	1.3	1.6

Считаме, че данните са от нормално разпределение. Може ли да се приеме, че двата метода дават в средно едни и същи резултати?

Задача 74. За да се сравни скоростта на четене при различни шрифтове, на две групи от по 50 човека е даден един и същи текст, но на първата бил написан с един шрифт, а на втората с друг. Регистрирано е времето, за което всеки участник е прочел текста. Резултатите са: за първата група – средно 7.88 минути и стандартно отклонение 1.73; за втората група – средно 8.48 и стандартно отклонение 2.12. Може ли да се твърди, че средното време за четене е различно за двата шрифта?

Задача 75. При проучване направено през 2007 година от 500 анкетирани 26 били вегетарианци. През 2019 година е направено ново проучване, според което 43 от 540 анкетирани отговорили, че са вегетарианци. Дават ли ни тези данни основание да твърдим, че процентът на вегетарианците през 2019 година се е увеличил спрямо 2007 година?

Задача 76. Генерирайте данни x_1, x_2, \dots, x_n от равномерно разпределение в интервала (5, 9). Намерете доверителен интервал за средното μ . Повторете 10000 пъти за $n = 20, 100, 500$. В каква част от случаите доверителният интервал съдържа 7?

Задача 77. Генерирайте данни x_1, x_2, \dots, x_n от равномерно разпределение в интервала (5, 9). Проверете хипотезата $\mu = 7$ срещу $\mu \neq 7$. Повторете 10000 пъти за $n = 20, 100, 500$. В каква част от случаите нулевата хипотеза не се отхвърля?

Задача 78. Генерирайте данни x_1, x_2, \dots, x_n от равномерно разпределение в интервала (5, 9). Намерете доверителен интервал за средното μ . Проверете хипотезата $\mu = 7$ срещу $\mu \neq 7$. Повторете 10000 пъти за $n = 20, 100, 500$. В каква част от случаите *доверителният интервал съдържа 7 и едновременно с това нулевата хипотеза не се отхвърля*?

Задача 79. Генерирайте данни x_1, x_2, \dots, x_n от $\mathcal{N}(5, 1)$ и y_1, y_2, \dots, y_n от $\mathcal{N}(5.2, 1)$. Проверете хипотезата за равенство на средните. Повторете 10000 пъти за $n = 20, 50, 100, 500, 1000$. В каква част от случаите нулевата хипотеза се отхвърля?

* * *

Задача 80. Резултатите от 180 хвърляния на зар са дадени в таблицата:

1	2	3	4	5	6
28	36	36	30	27	23

Може ли да се счита, че зарът е правилен?

Задача 81. Честотите на срещане на буквите в английския език са следните (в %):

E	T	A	O	I	N	S	R	H	D	other
12.02	9.10	8.12	7.68	7.31	6.95	6.28	6.02	5.92	4.32	26.28

В текст, състоящ се от 2004 букви, броят срещания на съответните букви е:

E	T	A	O	I	N	S	R	H	D	other
221	153	183	111	113	152	103	197	38	104	629

Може ли да се твърди, че текстът е на английски?

Задача 82. Според законите на Мендел, даден сорт грах може да има бели, розови или червени цветове, с вероятности съответно $1/4, 1/2, 1/4$. Направена е извадка от 564 растения от този сорт. От тях 141 цъфтели в бяло, 291 в розово и 132 в червено. В съгласие ли са тези резултати с теорията на Мендел?

Задача 83. В данните [pi2000](#) са първите 2000 цифри на числото π . Може ли да се счита, че всяка цифра се среща с една и съща вероятност?

Задача 84. Използвайте данните [survey](#) от пакета [MASS](#) и направете двумерна таблица на пушенето по пола. Проверете хипотезата за независимост между пушенето и пола.

Задача 85. Във файла [ManWomanEye.txt](#) има данни за цвета на очите на 204 семейни двойки. Има ли връзка между цвета на очите на мъжа и цвета на очите на жената?

Задача 86. Разгледайте данните [HairEyeColor](#). Има ли връзка между цвета на косата и цвета на очите?

Задача 87. В таблицата са представени данни за броя на пострадалите при катастрофи пътници, като са класифицирани по степен на нараняванията и по това дали пътникът е използвал препазен колан.

	Наранявания			
	Няма	Леки	Средни	Тежки
С колан	12813	647	359	42
Без колан	65963	4000	2642	303

Има ли връзка между използването на препазен колан и степента на нараняванията?

* * *

Задача 88. Във файла [bac.txt](#) има данни за съдържанието на алкохол в кръвта (грамаве алкохол на 100 мл кръв) в зависимост от броя на изпитите бутилки бира.

а) Постройте линеен модел. Напишете оцененото регресионно уравнение. Представете графично данните и построения линеен модел.

б) Интерпретирайте оценените коефициенти.

в) Имаме ли основание да твърдим, че има линейна връзка между броя на изпитите бутилки бира и съдържанието на алкохол в кръвта?

г) Може ли да се твърди, че при изпиването на още една бира съдържанието на алкохол в кръвта се увеличава средно с 0.02?

д) Намерете доверителен интервал за средното съдържание на алкохол в кръвта при 5 изпити бутилки бира.

е) За стойности на предиктора между 1 и 9, намерете доверителен интервал за средното съдържание на алкохол в кръвта и интервал за прогноза и ги илюстрирайте на графиката от а).

Задача 89. Генерирайте данни (x_i, y_i) , $i = 1, \dots, 50$, където:

x_i са случайни числа от равномерно разпределение в интервала $(1, 7)$;

ε_i са случайни числа от нормално разпределение с параметри $\mu = 0$ и σ ;

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

Разгледайте следните случаи за β_0 , β_1 , σ :

а) $\beta_0 = 2$, $\beta_1 = 1.5$, $\sigma = 2$.

б) $\beta_0 = 2$, $\beta_1 = 1.5$, $\sigma = 1$.

в) $\beta_0 = 2$, $\beta_1 = 0.17$, $\sigma = 1$.

Постройте линеен модел по данните (x_i, y_i) , $i = 1, \dots, 50$. Представете графично данните и построения линеен модел. Намерете R^2 и доверителен интервал за β_1 .

Задача 90. Генерирайте данни (x_i, y_i) , $i = 1, \dots, 50$, където:

x_i са случайни числа от равномерно разпределение в интервала $(1, 7)$;

ε_i са случайни числа от нормално разпределение с параметри $\mu = 0$ и $\sigma = 2.5$;

$y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$;

$\beta_0 = 2$, $\beta_1 = 1.1$.

Постройте линеен модел по данните (x_i, y_i) , $i = 1, \dots, 50$. Представете графично данните и построения линеен модел. Намерете R^2 и доверителен интервал за β_1 .

Задача 91. Във файла `satgpa.txt` има данни за резултатите на 1000 студенти. Разгледайте следните променливи:

`sat_v` резултат на Verbal SAT;
`sat_m` резултат на Math SAT;
`sat_sum` общ резултат на теста SAT;
`hs_gpa` успех от гимназията;
`fy_gpa` успех от първата година в колежа.

а) Постройте модел, който може да се използва за прогнозиране на успеха от първата година в колежа в зависимост от успеха от гимназията. Напишете оцененото регресионно уравнение. Направете подходяща графика.

б) Интерпретирайте коефициента пред предиктора. Може ли да се твърди, че има линейна връзка между двете променливи?

в) Прогнозирайте успеха от първата година в колежа на студент, чиито успех от гимназията е 3.5. Намерете доверителен интервал и интервал за прогноза.

г) Може ли да се получи по-добър модел, ако се включат и някои от останалите променливи?

Задача 92. Измерени са диаметъра (в инчове), височината (във футове) и обема (в кубични футове) на 31 черешови дървета. Диаметърът е измерен на височина 54 инча от земята. Данните са във файла `cherry.txt`.

а) Постройте модел, който може да се използва за прогнозиране на обема според диаметъра на дървото. Напишете оцененото регресионно уравнение. Направете подходяща графика.

б) Интерпретирайте коефициента пред диаметъра.

в) Постройте модел, който включва и височината. Интерпретирайте коефициента пред височината и коефициента пред диаметъра.

г) Намерете доверителен интервал за средния обем на дърво с диаметър 14 инча и височина 70 фута.

д) Постройте модел, в който участва диаметърът на квадрат. Намерете доверителен интервал за средния обем на дърво с диаметър 14 инча и височина 70 фута и сравнете с резултата от г).

Задача 93. Във файла `duke_forest.csv` има данни за 98 продадени къщи. Разгледайте следните променливи:

`price` цена (в долари);
`bed` брой спални;
`bath` брой бани;
`area` площ (в квадратни футове);
`year_built` година на построяване;
`lot` площ на двора на мястото (в акри).

Постройте модел, който може да се използва за прогнозиране на цената според характеристиките на къщата.

Задача 94. Генерирайте данни (x_i, y_i) , $i = 1, \dots, n$, където:

x_i са случайни числа от равномерно разпределение в интервала $(1, 10)$;

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$;

$\beta_0 = 2$, $\beta_1 = 5$.

Разгледайте следните случаи за ε_i :

а) ε_i са случайни числа от нормално разпределение с параметри $\mu = 0$, $\sigma = 5$;

б) ε_i са случайни числа от експоненциално разпределение с параметър $1/5$.

Постройте линеен модел по данните (x_i, y_i) , $i = 1, \dots, n$. Повторете 10000 пъти за $n = 30, 50, 100, 500$. За всеки от случаите и всяко n намерете:

- средното на $\hat{\beta}_1$ на базата на 10000 повторения;
- колко често доверителният интервал за β_1 съдържа истинската стойност;
- средната дължина на доверителния интервал за β_1 на базата на 10000 повторения.

* * *

* * *

Задача 95. За да се установи времето, необходимо на ученик да прочете даден текст, е направен тест с 30 ученици. Резултатите са следните (в минути):

25 29 18 29 22 20 27 24 20 29 18 20 31 25 21 24 24 21 18 24 24 29 25 24 27 22 25 22 27 25

Може ли да се твърди, че средното време, необходимо за прочитане на текста, е по-малко от 25 минути?

Задача 96. Разполагаме с данни за земетресенията в Южна Калифорния по ден от седмицата, за определен период от време (включени са земетресения с магнитуд поне 4.4 по Рихтер):

Ден от седмицата	пн	вт	ср	чт	пт	сб	нд
Брой земетресения	144	170	158	172	148	152	156

Дават ли ни тези данни основание да твърдим, че земетресение е равновероятно да се случи в кой да е ден от седмицата?

Задача 97. Извършен е следният експеримент с цел да се сравни добива от два сорта домати. Избрани са случайно по 36 растения от всеки сорт и е записан добива от всяко растение. Данните са във файла [tomato2.txt](#). Имаме ли основание да твърдим, че средният добив от втория сорт е по-голям?

Задача 98. За да се сравни надеждността на две марки електромотори, са тествани по 30 мотора от всяка марка. От първата марка 22 мотора преминали теста успешно, докато от втората успешно преминали 16 мотора. Може ли да приемем, че електромоторите от първата марка са по-надеждни?

Задача 99. За да се изследва доколко се различават цените в две книжарници, са избрани 73 книги и за всяка книга е записана цената в първата книжарница и цената във втората книжарница. Данните са във файла [books.txt](#). Имаме ли основание да твърдим, че в средно в първата книжарница цените са по-високи?

Задача 100. Според проучване проведено през 2003 година, 18% от използващите мобилни телефони са получавали рекламни съобщения по мобилните си телефони. През 2004 година е проведено ново изследване, според което 1100 от 5500 запитани потребители на мобилни телефони са получавали рекламни съобщения. Имаме ли основание да твърдим, че през 2004 година използващите мобилни телефони са получавали повече рекламни съобщения?

Задача 101. Използвайте данните [survey](#) от пакета [MASS](#) и намерете модел за прогнозиране на височината в зависимост от дължината на педята.

* * *