МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Пермский государственный национальный исследовательский университет»

Колледж профессионального образования

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания по практической работе

для студентов Колледжа профессионального образования

по специальностям 09.02.04 Информационные системы (по отраслям)

Утверждено на заседании ПЦК
Информационных технологий
Протокол № 9 от 23.05.2018
председатель __________Н.А. Серебрякова

Пермь 2018

Составители:

Ежова Марина Алексеевна – преподаватель ФГБОУ ВО ПГНИУ

Теория вероятностей и математическая статистика: методические указания по практической работе для студентов Колледжа профессионального образования специальности 09.02.04 Информационные системы (по отраслям) / сост. М.А. Ежова; Колледж проф. образ. ПГНИУ. – Пермь, 2018. – 36 с.

Методические указания «Теория вероятностей и математическая статистика» разработаны на основе требований Федерального государственного образовательного стандарта среднего профессионального образования по специальности 09.02.04 Информационные системы по отраслям для оказания помощи студентам специальности 09.02.04 Информационные системы (по отраслям) по дисциплине «Теория вероятностей и математическая статистика».

Предназначены для студентов Колледжа профессионального образования ПГНИУ специальности 09.02.04 Информационные системы (по отраслям) (СПО) всех форм обучения.

Оглавление

1. Комбинаторика

Теоретическая справка

Размещениями (без повторения) называются комбинации, составленные из **n**различных элементов по **m,** которые различаются либо составом элементов, либо их порядком. Число размещений из **n**различных элементов по **m**можно вычислить по формуле:

$$A_n^m = \frac{n!}{(n-m)!} = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-m+1)$$

Перестановками (без повторения) называются комбинации, состоящие из одних и тех же **n**различных элементов и отличающиеся друг от друга только порядком их расположения. Число перестановок из **n** различных элементов можно вычислить по формуле:

$$P_n = n!$$

Сочетаниями (без повторения) называются комбинации, составленные из **n**различных элементов по **m,** которые отличаются хотя бы одним элементом. Число сочетаний из **n**различных элементов по **m**можно вычислить по формуле:

$$C_n^m = \frac{n!}{m!\cdot(n-m)!}.$$

Правило суммы. Если некоторый объект**A** может быть выбран из совокупности объектов **m** способами, а другой объект **B** может выбран **n** способами, то выбрать либо **A**, либо **B**, можно (**m**+**n**) способами.

Правило произведения. Если объект**A** может быть выбран из совокупности объектов **m** способами, и после каждого такого выбора объект **B** можно выбрать **n** способами, то пара объектов (**A**, **B**) в указанном порядке, может быть выбрана **mn** способами.

Примеры решения задач

1. Вычислить A_9^3; C_8^4; $\frac{P_3 - P_2}{P_5}$.

Решение:

$A_9^3 = 9\cdot 8\cdot 7 = 504$.

$$C_8^4 = \frac{8!}{4!\cdot(8-4)!} = \frac{8!}{4!\cdot 4!} = \frac{5\cdot 6\cdot 7\cdot 8}{1\cdot 2\cdot 3\cdot 4} = 70.$$

$$\frac{P_3 - P_2}{P_5} = \frac{3!-2!}{5!} = \frac{1\cdot 2\cdot 3 - 1\cdot 2}{1\cdot 2\cdot 3\cdot 4\cdot 5} = \frac{6-2}{120} = \frac{4}{120} = \frac{1}{30}.$$

2. Имеется 20 карточек с буквами. Среди них поочерёдно выбираются 5. Сколько вариантов буквосочетаний можно составить?

Решение:

Так как повторения не допускаются и порядок выбора карточек важен, то используем формулу размещения:

$$A_{20}^5 = 20\cdot 19\cdot 18\cdot 17\cdot 16 = 1860480$$

Варианты заданий

1. Вычислить A_9^4; C_{12}^4.

2. Вычислить A_7^2; C_9^6.

3. Вычислить $\frac{P_6 - P_5}{P_4}$.

4. Вычислить A_8^4; C_{10}^4.

5. Вычислить A_7^3; C_7^6.

6. Вычислить $\frac{P_5 - P_3}{P_4}$.

7. Вычислить $A_7^2; C_9^6$.

8. Вычислить A_8^4; C_7^2.

9. Вычислить $\frac{P_6 - P_5}{P_2}$.

10. Вычислить $\frac{P_6 + P_5}{P_4}$.

11. В группе из 15 студентов выбирают старосту и представителя в студсовет. Сколько способов избрания на должности существуют?

12. За столом стоят семь стульев. Сколько способов рассадки гостей возможно?

13. В грузовой фуре 50 коробок. Из них таможенник должен проверить две. Подсчитать число версий осмотра.

14. В соревнованиях участвуют пять команд. Сколько различных способов заполнить итоговую турнирную таблицу от победителя до последнего места?

15. В столовой восемь видов салатов, пять – горячего и два – напитков. Сосчитать число вариантов обеда.

16. В группе из 20 студентов разыгрывают шесть билетов в кино. Сосчитать число способов раздать билеты.

17. Студент пытается вспомнить размер своей стипендии. Он знает, что это трехзначное число, а так же уверен, что там нет цифр 2, 7 и 0. Сколько вариантов чисел должен перебрать студент, если цифры не повторяются?

18. У девушки пятнадцать пар туфель, десять юбок и двенадцать блузок. Сколько нарядов может составить девушка?

19. Имеются кубики с буквами, из которых составлено слово «папаха». Сколько буквосочетаних можно составить из четырёх кубиков?

20. Юноши проходят медосмотр для военкомата друг за другом. Сколько вариантов упорядочить пятерых юношей?

2. Классическое определение вероятности

Теоретическая справка

Вероятность события А – отношение числа исходов, благоприятных появлению события **А**, к общему числу элементарных исходов.

$$p = \frac{m}{n}.$$

Пример решения задачи

В цехе работают 11 мужчин и 4 женщины. По табельным номерам наудачу выбраны семь человек, которые пойдут в отпуск. Найти вероятность того, что в отпуск пойдут 2 женщины.

Решение:

Всего в цехе работают 11 + 4 = 15 человек.

Пусть событие**А** – из 7 выбранных человек окажутся 2 женщины (5 мужчин, соответственно). Оно состоит из одновременного выполнения двух условий: выбрали 2 женщины из 4 и выбрали 5 мужчин из 11.

Пусть **n** – число всевозможных вариантов выбора 7 человек из 15:

$$n = C_{15}^{7} = \frac{15!}{7!\cdot(15-7)!} = \frac{15!}{7!\cdot 8!} = \frac{9\cdot 10\cdot 11\cdot 12\cdot 13\cdot 14\cdot 15}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7} = 6\,435$$

Путь $\mathbf{m_1}$ – число всевозможных вариантов выбора 2 женщин из 4, а $\mathbf{m_2}$ – число всевозможных вариантов выбора 5 мужчин из 11:

$$m_1 = C_4^2 = \frac{4!}{2!\cdot(4-2)!} = \frac{4!}{2!\cdot 2!} = \frac{3\cdot 4}{1\cdot 2} = 6$$

$$m_2 = C_{11}^{5} = \frac{11!}{5!\cdot(11-5)!} = \frac{11!}{5!\cdot 6!} = \frac{7\cdot 8\cdot 9\cdot 10\cdot 11}{1\cdot 2\cdot 3\cdot 4\cdot 5} = 462.$$

Тогда вероятность события**А** будет вычисляться как:

$$P(A) = \frac{m_1\cdot m_2}{n} = \frac{6\cdot 462}{6435} \approx 0.431.$$

Ответ: $P(A) \approx 0.431$.

Варианты заданий

1. В студенческой группе 4 девушки и 6 парней. Среди студентов разыгрываются 4 флаера в клуб. Найти вероятность того, что в клуб пойдут одни девушки.

2. Номера телефонов в некотором городе состоят из шести цифр. Какова вероятность того, что в наудачу выбранном номере все цифры четные.

3. Среди кандидатов на конференцию рассматривают 4 студентов с первого курса, 7 – со второго и 8 – с третьего. Из этого состава наудачу выбирают 5 человек. Найти вероятность того, что будут выбраны 2 второкурсника и 3 первокурсника.

4. В группе 12 студентов, среди которых 8 отличников. По списку выбрали 9 студентов. Найти вероятность того, что среди них пять отличников.

5. Студент выучил 5 вопросов из 35. Какова вероятность того, что из предложенных двух вопросов он знает ответ на один?

6. Устройство состоит из 6 элементов, из которых 3 нуждаются в ремонте. При запуске устройства включаются случайным образом 2 элемента. Найти вероятность того, что включенными окажутся неизношенные элементы.

7. Студент выучил 20 вопросов из 35. Какова вероятность того, что из предложенных трех вопросов он знает ответ на два.

8. Устройство состоит из пяти элементов, из которых два изношены. При запуске устройства включаются случайным образом два элемента. Найти вероятность того, что включенными окажутся неизношенные элементы.

9. Студент выучил половину из 20 вопросов. Какова вероятность того, что из предложенных четырёх вопросов он знает ответ на три?

10. Номера телефонов в некотором городе состоят из семи цифр. Какова вероятность того, что в наудачу выбранном номере все цифры нечетные?

3. Теоремы сложения и умножения вероятностей

Теоретическая справка

Теорема. Вероятность суммы двух независимых событий равна сумме их вероятностей.

$$P(A+B) = P(A) + P(B).$$

Теорема. Вероятность произведения двух независимых событий равна произведению их вероятностей.

$$P(A \cdot B) = P(A) \cdot P(B).$$

Пример решения задачи

Вероятности того, что в каждом из четырех магазинов имеется планшет нужной мощности, соответственно равны: 0.3; 0.45; 0.75 и 0.05. Найти вероятность того, что этот планшет будет куплен только в одном магазине.

Решение:

Пусть $\mathbf{A_i}$ – событие «в $\boldsymbol{i}$-ом магазине имеет нужный планшет». Обозначим вероятности этих собитий через p_i:

$$P(A_1) = p_1 = 0.3,$$
$$P(A_2) = p_2 = 0.45,$$
$$P(A_3) = p_3 = 0.75,$$
$$P(A_4) = p_4 = 0.05.$$

Вероятности обратных событий (в магазине нет планшета) соответственно равны:

$$P(\overline{A_1}) = q_1 = 0.7,$$
$$P(\overline{A_2}) = q_2 = 0.55,$$
$$P(\overline{A_3}) = q_3 = 0.25,$$
$$P(\overline{A_4}) = q_4 = 0.95.$$

Пусть $\mathbf{B_i}$ – событие «покупка произошла только в $\boldsymbol{i}$-ом магазине», т.е. в других магазинах планшета не было:

$$B_1 = A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cdot \overline{A_4},$$
$$B_2 = \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cdot \overline{A_4},$$
$$B_3 = \overline{A_1} \cdot \overline{A_2} \cdot A_3 \cdot \overline{A_4},$$
$$B_4 = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3} \cdot A_4.$$

Вероятности этих событий считаются следующим образом:

$$P(B_1) = P(A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cdot \overline{A_4}) = P(A_1) \cdot P(\overline{A_2}) \cdot P(\overline{A_3}) \cdot P(\overline{A_4}) = p_1 \cdot q_2 \cdot q_3 \cdot q_4 \approx 0.039,$$
$$P(B_2) = q_1 \cdot p_2 \cdot q_3 \cdot q_4 \approx 0.075,$$
$$P(B_3) = q_1 \cdot q_2 \cdot p_3 \cdot q_4 \approx 0.274,$$
$$P(B_4) = q_1 \cdot q_2 \cdot q_3 \cdot p_4 \approx 0.005.$$

Тогда событие**C** – «покупка совершена только в одном магазине» – состоит из того, что планшет куплен или в первом магазине, или во втором, или в третьем, или в четвёртом, т.е. $C = B_1 + B_2 + B_3 + B_4$.

Вероятность события **C**:

$$P(C) = P(B_1 + B_2 + B_3 + B_4) = P(B_1) + P(B_2) + P(B_3) + P(B_4) \approx 0.393.$$

Ответ: $P(B) \approx 0.393$.

Варианты заданий

1. Мужчина, посетивший торговый центр делает покупку с вероятностью 0.13, а женщина – с вероятностью 0.61. У прилавка одной из секций стоят 2 женщины и 1 мужчина. Они незнакомы друг с другом. Найти вероятность того, что только один покупатель сделает покупку.

2. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0.73; для второго – 0.82. Найти вероятность того, что будет два попадания в мишень, если каждому стрелку дается право на два выстрела.

3. Вероятности того, что в пяти магазинах электроники имеется телевизор нужной диагонали, соответственно равны: 0.9; 0.79; 0.85; 0.82 и 0.75. Найти вероятность того, что этого телевизора не будет ни в одном магазине.

4. Компьютер имеет три процессора и способен функционировать, если работают, по крайней мере, два из них. Вероятность выхода из строя 1-го процессора – 0.01; 2-го – 0.02; 3-го – 0.03. Какова вероятность выхода из строя компьютера?

5. Мужчина, посетивший торговый центр делает покупку с вероятностью 0.15, а женщина – с вероятностью 0.65. У прилавка одной из секций стоят одна женщина и двое мужчин. Они незнакомы друг с другом. Найти вероятность того, что двое покупателей сделают покупку.

6. Вероятность того, что в страховую компанию в течение года обратится с иском о возмещении ущерба 1-ый клиент, равна 0.108. Для второго эта вероятность равна 0.25; а для третьего – 0.62. Обращение клиентов – независимые события. Найти вероятность того, что в течение года обратятся два клиента.

7. Рабочий обслуживает три станка, работающих независимо друг от друга. Вероятность того, что в течение часа первый станок потребует внимания рабочего, равна 0.3; для второго эта вероятность равна 0.27; для третьего – 0.18. Какова вероятность того, что два станка не потребует внимания рабочего?

8. Вероятность того, что в четырех магазинах электроники имеется ультрабук нужной мощности, соответственно равны: 0.2; 0.32; 0.18 и 0.4. Найти вероятность того, что этого устройства не будет ни в одном магазине.

9. Компьютер имеет четыре ядра и способен функционировать, если работает хотя бы одно из них. Вероятность выхода из строя первого ядра – 0.15; второго – 0.1; третьего – 0.05; четвёртого – 0.2. Какова вероятность, что компьютер работает?

10. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0.7; для второго – 0.8. Найти вероятность того, что будет три попадания в мишень, если каждому стрелку дается право на два выстрела.

4. Формула полной вероятности. Формула Байеса

Теоретическая справка

Теорема. Вероятность события**A**, которое может наступить лишь при условий появления одного из попарно несовместных событий **B1**, **B2**, …, **Bn**, образующих полную группу, равна сумме произведения вероятностей, каждого из этих событий, на соответствующую условную вероятность события **A**.

$$P(A) = P(B_1) * P_{B_1}(A) + P(B_2) * P_{B_2}(A) + ... + P(B_n) * P_{B_n}(A).$$

Эта запись называется **формулой полной вероятности**.

Пусть событие**A** может наступить при условий появления одного из попарно несовместных событий **B1**, **B2**,…., **Bn**, образующих полную группу.

Из условия, что событие**A** уже произошло, переоценим вероятности гипотез:

$$P_A(B_i) = \frac{P(B_i) * P_{B_i}(A)}{P(A)}, \text{ где } 1 \le i \le n.$$

Эта запись называется **формулой Байеса**.

Пример решения задачи

Лекарство проверяется на полезность одним из двух экспертов. Вероятность того, что лекарство попадет к первому эксперту, равна 0.6; а ко второму – 0.4. Вероятность того, что полезное лекарство будет признано полезным первым экспертом, равна 0.85, а вторым – 0.9. Полезное лекарство при проверке было признано полезным. Найти вероятность того, что это лекарство проверил второй эксперт.

Решение:

Пусть событие**A** – лекарство было признано полезным. Оно состоит из суммы двух событий: лекарство попало к первому эксперту (гипотеза **B1**) и было признано полезным (вероятность – $P_{B_1}(A)$) или лекарство попало ко второму эксперту (гипотеза **B2**) и было признано полезным (вероятность $-P_{B_2}(A)$).

Пропишем вероятности всех событий:

$P(B_1) = 0.6$, $P(B_2) = 0.4$,

$P_{B_1}(A) = 0.85$, $P_{B_2}(A) = 0.9$.

Тогда вероятность события**A** вычисляется по формуле полной вероятности:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) = 0.6 \cdot 0.85 + 0.4 \cdot 0.9 = 0.87.$$

Ответ на задачу получим, используя формулу Байеса:

$$P_A(B_2) = \frac{P(B_2) \cdot P_{B_2}(A)}{P(A)} = \frac{0.4 \cdot 0.9}{0.87} \approx 0.414.$$

Ответ: $P_A(B_2) \approx 0.414$.

Варианты заданий

1. В специализированную больницу поступают в среднем 40% больных с заболеванием L, 25% - с заболеванием M, остальные с заболеванием K. Вероятность полного излечения болезни L равна 0.75, для болезни M и K эти вероятности соответственно равны 0.8 и 0.9. Больной, поступивший в больницу, был выписан здоровым. Найти вероятность того, что этот больной страдал заболеванием K.

2. В группу спортсменов – стрелков входят 5 новичков, 9 перворазрядников и 6 второразрядников. Вероятность попадания при одном выстреле для спортсменов следующие: 0.93; 0.82; 0.71. Наудачу выбранный стрелок попадает в мишень. Найти вероятность того, что в цель попал новичок или второразрядник.

3. Каждый день в магазин поступают овощи с трех продуктовых баз. Продукты с первой базы составляют 45%, со второй – 24%, а с третьей – 31% от всего объёма овощей. Вероятность срыва поставок соответственно равны: 0.02; 0.06; 0.03. Найти вероятность того, что в случайно выбранный день не будет срыва поставок.

4. На железнодорожном вокзале работают три билетные кассы. Вероятность того, что пассажир подойдет к первой кассе, равна $\frac{1}{3}$, ко второй – $\frac{2}{5}$ и к третьей – $\frac{4}{15}$. Вероятность того, что в нужном направлении будут билеты в кассах, равна соответственно $\frac{2}{3}$, $\frac{1}{4}$ и $\frac{1}{6}$. Пассажир купил билет. Найти вероятность того, что билет был приобретен во второй кассе.

5. Турист, блуждая по лесу, вышел на поляну, откуда ведут 5 тропинок. Вероятность выхода из леса в течение часа по каждой из этих тропинок равна соответственно 0.59; 0.27; 0.39; 0.41 и 0.12. Турист вышел из леса в течение часа. Найти вероятность того, что он вышел по второй или пятой тропинке.

6. В вычислительной лаборатории имеется одинаковое число компьютеров трех поколений. Вероятность сбоя компьютеров каждого типа соответственно равна 0.018; 0.016; 0.011. Наудачу выбранный компьютер во время работы не дал сбоя. Найти вероятность того, что это компьютер третьего типа.

7. На строительную площадку поступает цемент с четырёх заводов: с первого – 20%, со второго – 35%, с третьего – 25%, с четвёртого - 20% всего объёма. Стандартность цемента с первого завода составляет 98%, со второго – 97%, с третьего – 99%, а с четвёртого - 95%. Наудачу взятый мешок цемента оказался нестандартным. Найти вероятность того, что этот цемент произведен на втором заводе.

8. В жилом доме имеются квартиры трех типов планировок в соотношении 5:7:8. Вероятности того, что они не проданы, соответственно равны: 0.02; 0.07; 0.11. Найти вероятность того, что наудачу выбранная квартира окажется проданной.

9. Три охотника одновременно выстрелили в кабана. Позже обнаружилось, что кабан был убит только одной пулей. Найти вероятность того, что кабан был убит третьим охотником, если вероятности попадания у охотников следующие: 0.42; 0.63; 0.54.

10. В складе магазина электроприборов имеются холодильники трех комплектаций в соотношении 19:13:18. Вероятности продажи холодильника каждого типа составляют 0.15; 0.27 и 0.3, соответственно. Найти вероятность того, что проданный холодильник был первой комплектации.

5. Повторные испытания

Теоретическая справка

Вычислим вероятность того, что в **n**испытаниях событие**A** произойдёт **k**раз и не произойдёт **(n – k)** раз. Для этого можно использовать **формулу Бернулли**:

$$P_n(k) = C_n^k p^k q^{n-1} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Эта формула дает хорошие результаты, однако, при больших значениях **n**ее использование неудобно из-за громоздких вычислений. Условие применимости: $n < 10$.

Локальная теорема Лапласа. Если вероятность появления события**A** в каждом испытаний постоянна и отлична от 0 и 1, то вероятность того, что событие **A** в **n** испытаниях произойдет ровно **m** раз, приближенно равно:

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}}, \text{ где } x = \frac{k-np}{\sqrt{npq}}.$$

Обозначим $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}}$, тогда $P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x)$, где $x = \frac{k-np}{\sqrt{npq}}$, а для $\varphi(x)$ составлены таблицы значений, учитывая, что $\varphi(-x) = \varphi(x)$.

Условие применимости: $npq > 10$.

Интегральная теорема Лапласа. Если вероятность наступления события**A** в каждом испытаний постоянно и отлично от 0 и 1, то вероятность того, что событие **A** в **n** испытаний появиться от $\mathbf{k_1}$ до $\mathbf{k_2}$ раз, равно интегралу:

$$P_n(k_1, k_2) \approx \frac{1}{\sqrt{2\pi}} \cdot \int_{x_2}^{x_1} e^{-\frac{z^2}{2}} dz = \Phi(x_2) - \Phi(x_1), \text{ где } x_1 = \frac{k_1 - np}{\sqrt{npq}} \; x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

При решении задач используются таблицы значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$, при $x > 5$ $\Phi(x) = 0.5$, а при $x < 0$ $\Phi(-x) = \Phi(x)$.

Условие применимости: $npq > 10$.

Пример решения задачи

Для программиста определенной квалификации вероятность создать работающий сайт равна 0.8. За месяц он создал 300 сайтов. Найти вероятность того, что среди них от 190 до 250 работающих сайтов.

Решение:

Так как в задаче необходимо узнать вероятность попадания в промежуток $(190; 250)$, то при решении будем использовать интегральную теорему Лапласа:

$$P_n(m_1, m_2) \approx \Phi(x_2) - \Phi(x_1), \text{ где } x_1 = \frac{m_1 - np}{\sqrt{npq}}, x_2 = \frac{m_2 - np}{\sqrt{npq}}.$$

Определим значения переменных для данной задачи:

$n = 300$;

$p = 0.8 \rightarrow q = 1 - p = 0.2$;

$m_1 = 190$;

$m_2 = 250$.

Проверим, применима ли теорема для данной задачи:

$$npq = 300 \cdot 0{,}8 \cdot 0{,}2 = 48 > 10\,,$$

таким образом, теорема применима. Вычислим необходимые параметры:

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{190 - 300 \cdot 0{,}8}{\sqrt{48}} = -\frac{50}{\sqrt{48}} = -7.22\,,$$

$$x_2 = \frac{m_2 - np}{\sqrt{npq}} = \frac{250 - 300 \cdot 0{,}8}{\sqrt{48}} = \frac{10}{\sqrt{48}} = 1.44\,.$$

Тогда:

$$P_{300}(190{,}250) \approx \Phi(1.44) - \Phi(-7.22) = \Phi(1.44) + \Phi(7.22) = 0.4251 + 0.5 = 0.9251\,.$$

Ответ: $P_{300}(190{,}250) \approx 0.9251$.

Варианты заданий

1. Вероятность изготовления бракованного электросхемы равна 0.03. Определить вероятность того, что в изготовленной партии из 3000 штук окажется 5 бракованных схем.

2. Всхожесть семян данного растения составляет 87%. Найти вероятность того, что из 6 посеянных семян взойдут три.

3. В хлопковой ткани число длинных волокон составляет 85%. Какова вероятность того, что среди взятых наудачу 5 волокон длинных окажется три.

4. Книга издана тиражом в 60 000 экземпляров. Вероятность того, что в книге имеется дефект брошюровки, равна 0,1. Найти вероятность того, что тираж содержит 500 дефектных книг.

5. Известно, что 25% семян не имеют свойство всхожести. Какова вероятность того, что из 500 посаженных семян прорастут не менее 300 и не более 400 семян.

6. При оформлении сцены в осветительную сеть было включено 8 новых лампочек. Каждая лампочка в течение года перегорает с вероятностью 0.2. Найти вероятность того, что в течение года придется заменить не менее шести лампочек из поставленных.

7. Вероятность своевременного выполнения заказа по созданию сайта равна 0.75. Найти вероятность того, что из 160 заказов своевременно выполнят не менее 100 и не более 130.

8. Всхожесть семян экспериментального сорта пшеницы составляет 72%. Найти вероятность того, что из 9 посеянных семян взойдут четыре.

9. Вероятность изготовления бракованной компьютерной мыши равна 0.004. Определить вероятность того, что в изготовленной партии из 5 000 штук окажется 20 бракованных компьютерных мышей.

10. Книга «Программирование для чайников» издана тиражом в 10 000 экземпляров. Вероятность того, что в книге имеется дефект печати, равна 0.05. Найти вероятность того, что тираж содержит 200 дефектных книг.

6. Дискретные случайные величины

Теоретическая справка

Дискретной случайной величиной (ДСВ) называют случайную величину, которая принимает отдельные изолированные возможные значения с определенными вероятностями.

Функцией распределения случайной величины X называют функцию, задающую вероятность того, что случайная величина **X**, принимает значение меньшее, чем *x*:

$$F(x) = P(X < x)$$

Математическое ожидание равно сумме произведений всех возможных значений ДСВ на их вероятность:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots x_n p_n = \sum_i x_i p_i$$

Теорема. Математическое ожидание числа появлений события**A** в **n** независимых испытаниях равно произведению числа испытаний на вероятность появления этого события.

$$M(X) = np\ .$$

Дисперсией (рассеянием) **дискретной случайной величины** называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M\left([X - M(X)]^2\right).$$

Теорема. Дисперсия случайной величины равна разности между математическим ожиданием квадрата случайной величины **X** и квадратом ее математического ожидания:

$$D(X) = M\left(X^2\right) - [M(X)]^2.$$

Средним квадратическим отклонением случайной величины X(показатель разброса значений относительно среднего) называют квадратный корень из дисперсии:

$$\sigma(X) = \sqrt{D(X)}\ .$$

Пример решения задачи

Задана дискретная случайная величина:

x	-5	-1	1	3
p	0.18	0.3	0.42	0.1

Найти: а) числовые характеристики этой величины: математическое ожидание, дисперсию, среднее квадратическое отклонение; б) функцию распределения и построить ее график.

Решение:

а) Вычислим числовые характеристики:

$$M(X) = -5 \cdot 0.18 + (-1) \cdot 0.3 + 1 \cdot 0.42 + 3 \cdot 0.1 = -0.48\ .$$

$$M(X^2) = (-5)^2 \cdot 0.18 + (-1)^2 \cdot 0.3 + 1^2 \cdot 0.42 + 3^2 \cdot 0.1 = 6.12\ .$$

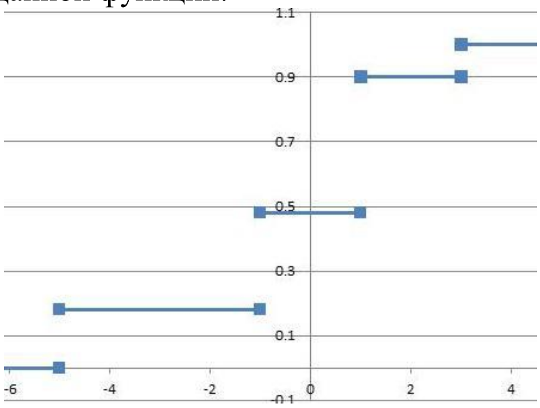
$$D(X) = M(X^2) - M^2(X) = 6.12 - (-0.48)^2 = 5.8896.$$

$$\sigma(X) = \sqrt{D(X)} \approx 2.427\ .$$

б) Найдём функцию распределения, последовательно складывая вероятности:

$$F(x) = \begin{cases} 0\,, & x \le -5; \\ 0.18\,, & -5 < x \le -1; \\ 0.48\,, & -1 < x \le 1; \\ 0.9\,, & 1 < x \le 3; \\ 1\,, & x > 3. \end{cases}$$

График этой кусочно-заданной функции:



Варианты заданий

1. Задан закон распределения дискретной случайной величины **X**:

x	-2	-1	2	4
p	0.4	0.3	0.2	0.1

Найти $F(x)$ и построить ее график.

2. Задан закон распределения дискретной случайной величины **X**:

x	-1	3	5	7
p	0.2	0.3	0.1	0.4

Найти числовые характеристики: математическое ожидание, дисперсию, среднее квадратическое отклонение.

3. Задан закон распределения дискретной случайной величины **X**:

x	1	3	6	9
p	0.2	0.3	0.4	0.1

Найти числовые характеристики случайной величины **X**: математическое ожидание, дисперсию и среднее квадратическое отклонение.

4. Задан закон распределения дискретной случайной величины **X**:

x	1	3	5	8
p	0.2	0.3	0.1	0.4

Найти $M(X)$, $D(X)$, $\sigma(X)$.

5. Задан закон распределения дискретной случайной величины **X**:

x	-1	0	2	5
p	0.12	0.08	0.33	0.47

Найти $F(x)$ и построить ее график.

6. Задан закон распределения дискретной случайной величины **X**:

x	-2	0	3	5
p	0.2	0.3	0.1	0.4

Найти числовые характеристики случайной величины **X**: $M(X)$, $D(X)$, $\sigma(X)$.

7. Задан закон распределения дискретной случайной величины **X**:

x	1	3	5	7
p	0.2	0.2	0.1	0.5

Найти $F(x)$ и построить ее график.

8. Задан закон распределения дискретной случайной величины **X**:

x	-2	-1	2	4
p	0.35	0.3	0.25	0.1

Найти $F(x)$ и построить ее график.

9. Задан закон распределения дискретной случайной величины **X**:

x	-1	3	5	7
p	0.2	0.3	0.15	0.35

Найти $M(X)$, $D(X), \sigma(X)$.

10. Задан закон распределения дискретной случайной величины **X**:

x	1	3	6	9
p	0.15	0.3	0.45	0.1

Найти числовые характеристики случайной величины **X**: математическое ожидание и среднее квадратическое отклонение.

7. Непрерывные случайные величины

Теоретическая справка

Непрерывной называют случайную величину, которая может принимать все значения из некоторого конечного или бесконечного промежутка. Число возможных значений, в этом случае, бесконечно.

Дифференциальной функцией распределения или **плотностью распределения вероятностей** называют первую производную интегральной функции распределения

$$f(x) = F\,'(x).$$

Математическим ожиданием непрерывной случайной величины, возможные значения которой принадлежат отрезку $[a\,;\,b]$, называется число, равное

$$M(x) = \int_a^b x f(x)dx.$$

Дисперсией непрерывной случайной величины называется математическое ожидание квадрата её отклонения:

$$D(x) = \int_a^b (x - M(x))^2 f(x)dx.$$

Специальная формула для вычисления дисперсии:

$$D(x) = \int_a^b x^2 f(x)dx - [M(x)]^2 == M(X^2) - M^2(X).$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{D(X)}.$$

Пример решения задачи

Задана непрерывная случайная величина:

$$f(x) = \begin{cases} 0, & x \le 0; \\ ax^2, & 0 \le x < 2; \\ 0, & x \ge 2. \end{cases}$$

Найти: коэффициент a; $M(X)$; $D(X)$; $\sigma(X)$; $F(x)$; $P(0.5 < X < 1.5)$. Построить графики функций $f(x)$ и $F(x)$.

Решение:

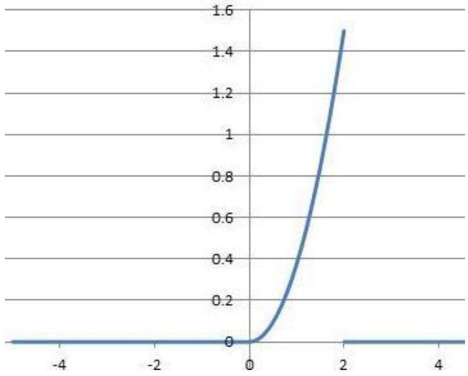
Исходя из свойства плотности вероятности $\int_{-\infty}^{+\infty} f(x)dx = 1$ получим:

$$\int_0^2 ax^2 dx = a \cdot \frac{x^3}{3}\bigg|_0^2 = \frac{a}{3}\left(2^3 - 0^3\right) = a \cdot \frac{8}{3} = 1.$$

Значит, $a = \frac{3}{8}$. Тогда

$$f(x) = \begin{cases} 0, & x \le 0; \\ \frac{3}{8}x^2, & 0 \le x \le 2; \\ 0, & x \ge 2. \end{cases}$$

График функции $f(x)$:



Интегральное распределение этой случайной величины будет, соответственно, равно:
При $x \le 0$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0dt = 0 .$$

При $0 \le x \le 2$

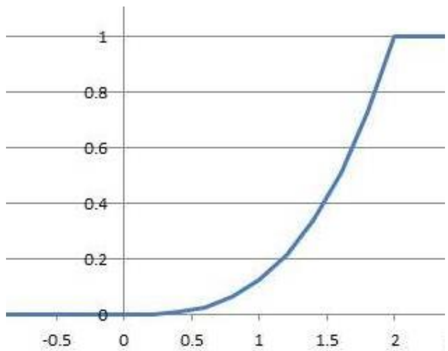
$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{x} \frac{3}{8}t^2 dt = 0 + \frac{3}{8}\frac{t^3}{3}\bigg|_0^x = \frac{x^3}{8} .$$

При $x \ge 2$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{2} \frac{3}{8}t^2 dt + \int_{2}^{x} 0dt = 0 + \frac{3}{8}\frac{t^3}{3}\bigg|_0^2 + 0 = \frac{8}{8} = 1.$$

$$F(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{x^3}{8}, & 0 \le x \le 2; \\ 1, & x \ge 2. \end{cases}$$

График этой функции:



Проверяем, действительно ли производная этой функции равна изначальной:

При $x \le 0$ $\quad F'(x) = (0)' = 0 .$

При $0 \le x \le 2$ $\quad F'(x) = \left(\dfrac{x^3}{8}\right)' = \dfrac{3x^2}{8} .$

При $x \ge 2$ $\quad F'(x) = (1)' = 0 .$

Таким образом, интегральное распределение случайной величины найдено правильно.

Вычислим числовые характеристики:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{2} x \cdot \frac{3}{8}x^2 dx + \int_{2}^{+\infty} 0dx = \int_{0}^{2} \frac{3}{8}x^3 dx = \frac{3}{8} \cdot \frac{x^4}{4}\bigg|_0^2 =$$

$$= \frac{3}{32}\left(2^4 - 0^4\right) = \frac{3 \cdot 16}{32} = 1.5\,.$$

$$M(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{2} x^2 \cdot \frac{3}{8} x^2 dx + \int_{2}^{+\infty} 0dx = \int_{0}^{2} \frac{3}{8} x^4 dx = \frac{3}{8} \cdot \left.\frac{x^5}{5}\right|_0^2 =$$

$$= \frac{3}{40}\left(2^5 - 0^5\right) = \frac{3 \cdot 32}{40} = 2.4\,.$$

$$D(X) = M(X^2) - M^2(X) = 2.4 - (1.5)^2 = 0.15\,.$$

$$\sigma(X) = \sqrt{D(X)} \approx 0.387\,.$$

Найдём вероятность попадания в интервал (0.5; 1.5):

$$P(0.5 < X < 1.5) = F(1.5) - F(0.5) = \frac{(1.5)^3}{8} - \frac{(0.5)^3}{8} = 0.40625\,.$$

Ответ: $M(X) = 1.5$; $D(X) = 0.15$; $\sigma(X) \approx 0.387$; $P(0.5 < X < 1.5) \approx 0.41$

Варианты заданий

1. Случайная величина задана плотностью распределения:

$$f(x)=\begin{cases}0, & x\le 0;\\ ax^2, & 0\le x\le 1;\\ 0, & x\ge 1.\end{cases}$$

Найти коэффициент a, $F(x)$ и построить ее график.

2. Случайная величина задана плотностью распределения:

$$f(x)=\begin{cases}0, & x\le 0;\\ 3x^2, & 0\le x\le 1;\\ 0, & x\ge 1.\end{cases}$$

Найти среднее квадратическое отклонение данной случайной величины.

3. Непрерывная случайная величина задана законом распределения:

$$F(x)=\begin{cases}0, & x\le 1.5;\\ 2x-3, & 1.5\le x\le 2;\\ 1, & x\ge 2.\end{cases}$$

Найти функцию плотности распределения вероятностей и построить графики этих функций.

4. Случайная величина задана плотностью распределения:

$$f(x)=\begin{cases}0, & x\le -1;\\ Cx^4, & -1\le x\le 1;\\ 0, & x\ge 1.\end{cases}$$

Найти коэффициент C и дисперсию данной случайной величины.

5. Случайная величина задана плотностью распределения:

$$f(x)=\begin{cases}0, & x\le 0;\\ 3x^2, & 0\le x\le 1;\\ 0, & x\ge 1.\end{cases}$$

Найти $P(0.25<X<0.75)$ и построить график функции $F(x)$.

6. Плотность распределения непрерывной случайной величины имеет вид:

$$f(x)=\begin{cases}0, & x\notin[0;\,2],\\ Cx^2, & x\in[0;\,2].\end{cases}$$

Определить константу C, построить функцию распределения $F(x)$ и вычислить вероятность $P(-1<X<1)$.

7. Интегральная функция распределения имеет вид:

$$F(x)=\begin{cases} 0, & x \le 1; \\ x^2-2x+1, & 1 \le x \le 2; \\ 1, & x \ge 2. \end{cases}$$

Найти $M(X)$ и $D(X)$.

8. Плотность распределения непрерывной случайной величины имеет вид:

$$f(x)=\begin{cases} 0, & x \le 2; \\ ax^2, & 2 \le x \le 3; \\ 0, & x \ge 3. \end{cases}$$

Найти коэффициент a, функцию $F(x)$ и вычислить $M(X)$.

9. Интегральная функция распределения случайной величины имеет вид:

$$F(x)=\begin{cases} 0, & x \le 0; \\ \dfrac{x^2}{2}-\dfrac{x^3}{8}, & 0 \le x \le 2; \\ 1, & x \ge 2. \end{cases}$$

Найти функцию $f(x)$, вычислить $M(X)$ и $D(X)$.

10. Плотность распределения непрерывной случайной величины имеет вид:

$$f(x)=\begin{cases} 0, & x \le 0; \\ Cx^2+3, & 0 \le x \le 3; \\ 0, & x \ge 3. \end{cases}$$

Найти коэффициентC, вычислить $P(1 < X < 2)$.

8. Вариационные ряды и их числовые характеристики

Теоретическая справка

Пусть изучается некоторая случайная величина **X**. С этой целью проведено **n** независимых испытаний, в результате которых получены значения: $x_1, x_2, \dots, x_n$.

Совокупность $x_1, x_2, \dots, x_n$ называется **статистическим рядом**.

Совокупность $x_1, x_2, \dots, x_k$, в которой значения расположены в порядке строгого возрастания и без повторения, называется **вариационным рядом**.

Значение x_i – **варианта**. Число наблюдений, соответствующих конкретной варианте, - её **частота**.

Число $w_i = \dfrac{n_i}{n}$ **- относительная частота (частность)**.

Статистическое распределение выборки – соответствие между вариантами вариационного ряда и их частотами или относительными частотами.

Полигон частот (частностей) – ломанная, соединяющая точки $(x_i; n_i)$ или $(x_i; w_i)$, соответственно.

Выборочное среднее – среднее арифметическое всех полученных значений случайной величины.

$$\bar{x} = M^*(X) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \; .$$

Выборочная дисперсия – среднее арифметическое квадратов отклонения полученных значений от выборочного среднего.

$$D^*(X) = D_X^* = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \; .$$

Для выборочной дисперсии справедлив аналог формулы для теоретической дисперсии:

$$D_X^* = \overline{x^2} - \bar{x}^2 \; .$$

Можно доказать, что такая оценка состоятельная и смещённая. В связи с этим в качестве оценки теоретической дисперсии берут **исправленную выборочную дисперсию**, вычисляемую по формуле:

$$S^2 = \frac{n}{n-1} D_X^* \; .$$

Эта оценка состоятельная и несмещённая.

В качестве оценки СКО берут **выборочное СКО**:

$$\sigma^*(X) = \sigma_X^* = \sqrt{D_X^*} \; .$$

и **«исправленное» выборочное СКО**:

$$S = \sqrt{S^2} \; .$$

Варианты заданий

1. Вычислить среднее арифметическое, выборочную дисперсию, выборочное среднее квадратическое отклонение по данному распределению:

x_i	31	35	37	39	40
n_i	20	25	25	20	10

2. По данному распределению найти: 1) среднее арифметическое; 2) выборочную дисперсию.

x_i	-5	-2	0	1	3
n_i	2	4	5	6	4

3. Построить полигон частот и частностей ряда распределения:

x_i	1	3	5	7	9
n_i	10	15	30	33	12

4. Найти выборочные $\bar{x}$ и D_X^* случайной величины **X**:

x_i	3.5	3.6	3.7	3.8	3.9	4
n_i	16	20	25	15	13	11

5. Вычислить относительные частоты и построить полигон частот по данному распределению:

x_i	6	10	12	14	15
n_i	15	20	25	20	10

6. По данному распределению найти выборочные $\bar{x}$ и D_X^* случайной величины **X**.

x_i	-5.5	-2.5	0	2.5	4.5
n_i	7	9	15	11	9

7. Построить полигон частот и частностей ряда распределения:

x_i	-1	2	5	8	11
n_i	10	15	30	23	12

8. Найти выборочное среднее квадратическое отклонение случайной величины **X**:

x_i	2.5	2.6	2.7	2.8	2.9	3
n_i	16	20	25	15	13	11

9. По данному распределению найти выборочные $\bar{x}$ и D_X^* случайной величины **X**.

x_i	-1	2	3	7	11
n_i	5	10	25	18	7

10. Найти выборочное среднее квадратическое отклонение случайной величины **X**:

x_i	1.1	1.2	1.3	1.4	1.5	1.6
n_i	6	10	15	5	3	1

9. Статистическая оценка параметров распределения генеральной совокупности

Теоретическая справка

Оценка неизвестного параметра Θ числом Θ^* - точная оценка. При большом числе испытаний она близка к оцениваемому параметру. Выясним, насколько велика вероятность **γ** того, что Θ^* мало отличается от Θ, т.е. вероятность неравенства:

$$\left|\Theta^* - \Theta\right| < \delta,$$

иначе:

$$\Theta^* - \delta < \Theta < \Theta^* + \delta,$$

где **δ – точность оценки**. Вероятность **γ** называется **доверительной вероятностью (надёжностью)**, а интервал $\left(\Theta^* - \delta; \Theta^* + \delta\right)$ - **доверительным интервалом** или интервальной оценкой параметра Θ.

1) Доверительный интервал для математического ожидания M(X) нормального распределения при известном*σ*.

Пусть случайная величина **X** имеет нормальное распределение, его СКО равно *σ*. Проведено **n** независимых испытаний, по результатам которых найдено выборочное среднее $\overline{x_в}$ - статистическая оценка неизвестного математического ожидания **a**. Для нахождения доверительного интервала используем функцию Лапласа:

$$\Phi(t) = \frac{\gamma}{2} \quad => \quad \delta = \frac{\sigma \cdot t}{\sqrt{n}} \quad => \quad \left(\overline{x_в} - \delta; \overline{x_в} + \delta\right).$$

2) Доверительный интервал для математического ожидания M(X) нормального распределения при неизвестном *σ*.

Пусть случайная величина **X** имеет нормальное распределение. Проведено **n** независимых испытаний, по результатам которых найдено выборочное среднее $\overline{x_в}$ и «исправленное» СКО *S*. Для нахождения доверительного интервала используем дополнительную переменную, имеющую распределение Стьюдента:

$$t_\gamma = f(\gamma, n) \quad => \quad \delta = \frac{S \cdot t_\gamma}{\sqrt{n}} \quad => \quad \left(\overline{x_в} - \delta; \overline{x_в} + \delta\right).$$

Составлены специальные таблицы, по которым можно найти число t_γ по значениям γ и **n**.

3) Доверительный интервал для СКО *σ*(X) нормального распределения.

Пусть случайная величина X имеет нормальное распределение. Проведено **n** независимых испытаний, по результатам которых найдено «исправленное» СКО *S* – статистическая оценка неизвестного СКО $\sigma(X)$.

Для нахождения доверительного интервала используем дополнительную переменную, значение которой берётся из специально составленных таблиц, исходя из чисел γ и **n**:

$$q = f(\gamma, n).$$

а) q< 1, то $\left(S \cdot (1-q); S \cdot (1+q)\right)$.

б) q ≥ 1, то $\left(0; S \cdot (1+q)\right)$.

Пример решения задачи

По данным n независимых равноточных измерений некоторой физической величины найдены среднеарифметическое результатов измерений $\overline{x_в}$ и генеральное среднее квадратичное отклонение $\sigma(X)$. Оценить истинное значение измеряемой величины с надежностью γ, если $n = 36$, $\overline{x_в} = 26.2$, σ = 6, γ = 0.99.

Решение:

Воспользуемся известным соотношением между функцией Лапласа и надёжностью:

$$\Phi(t) = \frac{\gamma}{2}.$$

Отсюда найдём t: $\Phi(t) = 0.495$, t = 2.58.

Теперь вычислим δ: $\delta = \frac{\sigma \cdot t}{\sqrt{n}} = \frac{6 \cdot 2.58}{6} = 2.58$.

Таким образом, доверительный интервал ищем по формуле:

$$\left(\overline{x_в} - \delta\, ;\, \overline{x_в} + \delta\right).$$

В частности

$$\left(26.2 - 2.58\, ;\;\; 26.2 + 2.58\right)$$

или

$$\left(23.62\, ;\, 28.78\right).$$

Ответ: $\overline{x_в} \in \left(23.62;\, 28.78\right)$.

Варианты заданий

1. Найти доверительный интервал для оценки математического ожидания нормального распределения с надежностью 0.95, зная $\overline{x_в}$ = 75,17, n = 49, σ = 6.

2. По данным **n** независимых равноточных измерений некоторой физической величины найдены среднеарифметическое результатов измерений $\overline{x_в}$ и генеральное среднее квадратичное отклонение σ. Оценить истинное значение измеряемой величины с надежностью γ, если n= 25, $\overline{x_в}$ = 14, σ = 5, γ = 0.95.

3. По данным **n** независимых равноточных измерений некоторой физической величины найдено «исправленное» среднее квадратичное отклонение S. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью γ, если n= 16, S = 1, γ= 0.99.

4. Найти доверительный интервал для оценки математического ожидания нормального распределения с надежностью 0.95, зная $\overline{x_в}$ = 75,1; n = 36; σ = 6.

5. По данным **n** независимых равноточных измерений некоторой физической величины найдены среднеарифметическое результатов измерений $\overline{x_в}$ и «исправленное» среднее квадратичное отклонение S. Оценить истинное значение измеряемой величины с надежностью γ, если n= 9, $\overline{x_в} = 30.1$, S = 6, γ = 0.99.

6. По данным **n** независимых равноточных измерений некоторой физической величины найдены среднеарифметическое результатов измерений $\overline{x_в}$ и генеральное среднее квадратичное отклонение σ. Оценить истинное значение измеряемой величины с надежностью γ, если n= 25, $\overline{x_в} = 16.8$, σ = 5, γ = 0.9.

7. Найти доверительный интервал для оценки математического ожидания нормального распределения с надежностью 0.95, зная $\overline{x_в}$ = 72.15; n = 36; σ = 4.

8. По данным **n** независимых равноточных измерений некоторой физической величины найдены среднеарифметическое результатов измерений $\overline{x_в}$ и «исправленное» среднее квадратичное отклонение S. Оценить истинное значение измеряемой величины с надежностью γ, если n= 25, $\overline{x_в}$ = 14, S= 5, γ = 0.95 .

9. По данным **n** независимых равноточных измерений некоторой физической величины найдены среднеарифметическое результатов измерений $\overline{x_в}$ и генеральное среднее квадратичное отклонение σ. Оценить истинное значение измеряемой величины с надежностью γ, если n= 16, $\overline{x_в}$ =10.2 , σ = 4, γ = 0.999.

10. По данным **n** независимых равноточных измерений некоторой физической величины найдено «исправленное» среднее квадратичное отклонение S. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью γ, если n= 50, S = 14, γ = 0.95.

10. Проверка статистических гипотез

Теоретическая справка

Числа n_i, полученные в результате эксперимента по измерению случайной величины, называются **экспериментальными частотами**.

Теоретическая частота $n_i^{'}$ - то число раз, которое случайная величина должна была попасть в ***i***-ый интервал, если она действительно распределена по нормальному закону.

При вычислении используют следующую таблицу:

№	x_i	x_{i+1}	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	P_i	$n_i^{'}$
…	…	…	…	…	…	…	…	…

и формулы:

$$z_i = \frac{x_i - \bar{x}}{\sigma_X^*} \qquad P_i = \Phi(z_{i+1}) - \Phi(z_i) \qquad n_i^{'} = P_i \cdot n.$$

Пусть в результате исследования найдены экспериментальные частоты n_iи в предположении о нормальном распределении вычислены $n_i^{'}$. Проверим гипотезу о нормальном распределении случайной величины. Для этого используем специальную величину – **критерий согласия** $\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$. По данной формуле вычисляется $\chi^2_{набл}$.

Составлены специальные таблицы, в которых по данному уровню значимости α и числу степеней свободы $c = k - 3$, где k – число различных значений x, определяют **критическую точку** $\chi^2_{кр}$.

Если $\chi^2_{кр} \geq \chi^2_{надл}$, т.е. $\chi^2_{набл}$ попадает в **область принятия гипотезы**, то гипотеза о нормальном распределении случайной величины принимается.

Если $\chi^2_{кр} \leq \chi^2_{надл}$, т.е. $\chi^2_{набл}$ попадает в **критическую область**, то гипотеза о нормальном распределении случайной величины не принимается.

Пример решения задачи

Используя критерий Пирсона, при уровне значимости 0.025 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности **X** с эмпирическим распределением выборки объема n = 50.

n_i	2	17	13	10	8
n'_i	6	14	18	7	5

Решение:

Определим основные параметры задачи:

α = 0.025,

n = 50,

k= 5 (число строк),

k – 3 = 2 (число степеней свободы),

Тогда $\chi^2_{кр} = 7.4$.

Составим таблицу и вычислим её элементы:

№	n_i	n'_i	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	2	6	16	2.67
2	17	14	9	0.64
3	13	18	25	1.39
4	10	7	9	1.29
5	8	5	9	1.80
Итого:	50	50		7.78

$\chi^2_{набл} = 7.78$.

Так как $\chi^2_{набл} > \chi^2_{кр}$, то гипотеза о нормальном распределении генеральной совокупности не принимается.

Ответ: гипотеза не принимается.

Варианты заданий

Используя критерий Пирсона, при уровне значимости 0.05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности **X** с эмпирическим распределением выборки объема **n** = 50.

1.

n_i	5	10	20	10	5
n'_i	6	14	18	7	5

$\alpha = 0.05$

2.

n_i	10	10	10	10	10
n'_i	6	14	18	7	5

$\alpha = 0.025$

3.

n_i	9	11	19	4	7
n'_i	6	14	18	7	5

$\alpha = 0.01$

4.

n_i	2	17	13	10	8
n'_i	6	14	18	7	5

$\alpha = 0.05$

5.

n_i	2	11	18	11	8
n'_i	6	14	18	7	5

$\alpha = 0.025$

6.

n_i	6	10	15	11	8
n'_i	6	14	18	7	5

$\alpha = 0.01$

7.

n_i	6	10	16	11	7
n'_i	6	14	18	7	5

$\alpha = 0.05$

8.

n_i	5	10	17	11	7
n'_i	6	14	18	7	5

$\alpha = 0.025$

9.

n_i	15	6	17	5	7
n'_i	6	14	18	7	5

$\alpha = 0.01$

10.

n_i	6	8	9	16	11
n'_i	6	14	18	7	5

$\alpha = 0.05$

11. Корреляционный анализ

Теоретическая справка

Две величины связаны **функциональной зависимостью**, если зная значение одной из них можно точно указать значение другой.

Статистическая зависимость: зная значение одной величины, можно указать закон распределения второй величины, зависящий от того, какое значение приняла первая величина.

Условным средним $\overline{y_x}$ называют среднее арифметическое значений величины **Y**, соответствующих значению **x** величины **X**:

$$\overline{y_x} = \frac{y_1 + y_1 + \dots y_m}{m}.$$

Корреляционная зависимость Y от X – функциональная зависимость условного среднего $\overline{y_x}$ от x.

Уравнение $\overline{y_x} = f(x)$ - **уравнение регрессииY на X**.

Функция $f(x)$ – **регрессияY на X**, график этой функции – **линия регрессии Y на X**.

Пусть случайные величины **X** и **Y** связаны линейной корреляционной зависимостью, т.е. обе линии регрессии прямые.

Проведено **n** независимых испытаний, в результате которых получено **n** пар значений: $(x_1; y_1)$, … ,$(x_n;\ y_n)$. Считаем, что каждому значению x_i и соответствующее ему значение y_iвстречаются по одному разу. Будем искать выборочное уравнение регрессии **Y** на **X** в виде:

$$\overline{Y_X} = \rho_{yx} x + b,$$

где ρ_{yx} - коэффициент регрессии **Y**на **X**. Таким образом, нужно найти параметры ρ_{yx} и ***b*** так, чтобы прямая как можно ближе располагалась к экспериментальным точкам. В результате которого получим:

$$\rho_{yx} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\left(\sigma_X^*\right)^2} \qquad \text{и} \qquad b = \overline{y} - \rho_{yx} \cdot \overline{x}.$$

Уравнение прямой линии регрессии **Y** на ***X*** будет выглядеть так: $\overline{Y_X} - \overline{y} = \rho_{yx}\left(x - \overline{x}\right)$.

Выборочный коэффициент корреляции – число, равное $r^* = \dfrac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\sigma_X^* \cdot \sigma_Y^*}$.

Тогда коэффициент регрессии вычисляется так: $\rho_{yx} = \dfrac{\sigma_Y^*}{\sigma_X^*} r^*$.

Если получено много данных, среди которых есть повторяющиеся, то их группируют и записывают в виде **корреляционной таблицы**. Если корреляционная таблица задана равноотстоящими вариантами с шагами $h_x = x_{i+1} - x_i$ и $h_y = y_{j+1} - y_j$, то переходят к условным вариантам: $u_i = \dfrac{x_i - C_x}{h_x}$ и $v_j = \dfrac{y_j - C_y}{h_y}$, где $(C_x;\ C_y)$ – **ложные нули**, расположенные примерно в середине таблице (нередко эта точка имеет наибольшую частоту). Тогда используем формулы:

$$\overline{x} = h_x \overline{u} + C_x, \quad \overline{y} = h_y \overline{v} + C_y, \qquad r^* = \frac{\overline{u \cdot v} - \overline{u} \cdot \overline{v}}{\sigma_u \cdot \sigma_v}, \qquad \sigma_X^* = h_x \cdot \sigma_u, \quad \sigma_Y^* = h_y \cdot \sigma_v.$$

Так как «ложный нуль» делит таблицу на 4 части, то такой метод называется «***методом четырёх полей***».

Пример решения задачи

Найти выборочное уравнение прямой линии регрессии **Y** на **X** по данным, приведенным в корреляционной таблице.

Y\X	5	8	11	14	17
50					2
55				3	1
60		1	11		
65		3			
70	3	1			

Решение:

Укажем исходные параметры:

n = 25,

h_x = 3 (шаг по переменной x),

h_y = 5 (шаг по переменной y).

Дополним таблицу следующими элементами:

Y\X	5	8	11	14	17	n_y
50					2	2
55				3	1	4
60		1	11			12
65		3				3
70	3	1				4
n_x	3	5	11	3	3	25

Строка n_x – суммарные частоты по переменной x (т.е. суммируем частоты по столбцам).
Столбец n_y – суммарные частоты по переменной y (т.е. суммируем частоты по строкам).

В качестве «ложного нуля» выберем ячейку с самой большой частотой (в данном случае 11). Тогда:

$$C_x = 11 \text{ и } C_y = 60.$$

Вычислим значения новых переменных по формулам:

$$U_i = \frac{x_i - C_x}{h_x} \text{ и } V_i = \frac{y_i - C_y}{h_y}.$$

Для этих переменных: $n_u = n_x$, $n_v = n_y$.

Теперь таблица выглядит так:

	U	-2	-1	0	1	2	n_y
V	**Y\X**	5	8	11	14	17	
-2	50					2	2
-1	55				3	1	4
0	60		1	11			12
1	65		3				3
2	70	3	1				4
n_x		3	5	11	3	3	25

Приступим к вычислениям:

$$\overline{U} = \frac{-2\cdot 3 + (-1)\cdot 5 + 0\cdot 11 + 1\cdot 3 + 2\cdot 3}{25} = -0.08;$$

$$\overline{V}=\frac{-2\cdot 2+(-1)\cdot 4+0\cdot 12+1\cdot 3+2\cdot 4}{25}=0.12;$$

$$\overline{U^2}=\frac{(-2)^2\cdot 3+(-1)^2\cdot 5+0^2\cdot 11+1^2\cdot 3+2^2\cdot 3}{25}=1.28;$$

$$\overline{V^2}=\frac{(-2)^2\cdot 2+(-1)^2\cdot 4+0^2\cdot 12+1^2\cdot 3+2^2\cdot 4}{25}=1.24;$$

$$\overline{UV}=\frac{-2\cdot 2\cdot 3+(-1)\cdot 0\cdot 1+(-1)\cdot 1\cdot 3+(-1)\cdot 2\cdot 1+0\cdot 0\cdot 11+1\cdot(-1)\cdot 3+2\cdot(-2)\cdot 2+2\cdot(-1)\cdot 1}{25}=-1.2$$

Теперь можно узнать следующие показатели:

$$\overline{X}=\overline{U}\cdot h_x+C_x=10.76; \qquad \overline{Y}=\overline{V}\cdot h_y+C_y=60.6.$$

Продолжим вычислять параметры для нахождения уравнения прямой линии регрессии:

$$\sigma^*(U)=\sqrt{\overline{U^2}-\left(\overline{U}\right)^2}\approx 1.129; \qquad \sigma^*(V)=\sqrt{\overline{V^2}-\left(\overline{V}\right)^2}\approx 1.107;$$

$$\sigma^*(X)=\sigma^*(U)\cdot h_x\approx 3.386; \qquad \sigma^*(Y)=\sigma^*(V)\cdot h_y\approx 5.535;$$

$$r^*=\frac{\overline{UV}-\overline{U}\cdot\overline{V}}{\sigma^*(U)\cdot\sigma^*(V)}\approx -0.953; \qquad \rho_{YX}=\frac{\sigma^*(V)}{\sigma^*(U)}r^*\approx -1.558.$$

Составим уравнение:

$$Y_X-\overline{Y}=\rho_{YX}\cdot\left(x-\overline{X}\right)$$

$$y-60.6=-1.558\cdot(x-10.76).$$

Приведём уравнение к стандартному виду:

$$y=-1.558x+77.362.$$

Ответ:

$y=-1.558x+77.362$.

Варианты заданий

Результаты наблюдений переменных **X** и **Y** приведены в таблице. Найти выборочный коэффициент корреляции.

1.

Y\X	30	35	40	45	50
170					2
175				2	1
180			15		
185		3			
190	2				

2.

Y\X	33	36	39	42	45
220					1
225				3	2
230			11		
235		4	1		
240	3				

3.

Y\X	25	29	33	37	41
210					1
215				5	1
220			11		
225		4			
230	3				

4.

Y\X	40	50	60	70	80
150	2				
155	3	4			
160		1	6		
165			3	3	1
170					2

5.

Y\X	35	45	55	65	75
150	3				
155		5	1		
160		1	5	1	
165			1	5	
170					3

6.

Y\X	12	17	22	27	32
150	3	2			
155		3	1		
160			8		
165			1	4	
170					3

7.

Y\X	20	25	30	35	40
150	2	3			
155		3			
160			9		
165			1	2	1
170					3

8.

Y\X	15	20	25	30	25
50				3	2
55				2	2
60		2	8		
65		2			
70	3	1			

9.

Y\X	5	8	11	14	17
50					2
55				3	1
60		1	11		
65		3			
70	3	1			

10.

Y\X	5	8	11	14	17
70	1	2			
80		3			
90			11		
100				2	2
110					4