Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Удмуртский государственный университет»

Институт математики, информационных технологий и физики

Кафедра математического анализа

ОПИСАТЕЛЬНАЯ СТАТИСТИКА И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Учебно-методическое пособие

Ижевск
2021

УДК 519.22(075.8)
ББК 22.172я73
О-611

Рекомендовано к изданию Учебно-методическим советом УдГУ

Рецензент д. т. н., проф. ФГБОУ ВО «ИжГТУ им. М. Т. Калашникова» И. Г. Русяк

Авторы-составители: А. Е. Ирисов, Ю. М. Сметанин

О-611 Описательная статистика и проверка статистических гипотез: учебно-методич. пособие / авторы-составители А. Е. Ирисов, Ю. М. Сметанин. Ижевск: Изд. центр «Удмуртский университет», 2021. — 58 с.

ISBN 978-5-4312-0889-8

В учебно-методическом пособии приведены краткие теоретические сведения, изложена методика выполнения работ по математической статистике и даны примеры решений задач, приведены индивидуальные задания типовых рассчетов.

Данное пособие предназначено для студентов бакалавриата всех нематематических специальностей, а также для магистрантов, изучающих курсы, связанные с математическим моделированием в Удмуртском государственном университете.

ISBN 978-5-4312-0889-8

9 785431 208898

УДК 519.22(075.8)
ББК 22.172я73

ВВЕДЕНИЕ

Данное учебно-методическое пособие предназначено для методического обеспечения практических занятий и самостоятельной работы в рамках курсов по теории вероятностей и математической статистике, изучаемых студентами, обучающимися по направлениям 19.03.01 Биотехнология, 21.03.01 Нефтегазовое дело, по специальности 38.05.01 Экономическая безопасность.

Содержание пособия соответствует рабочим программам по указанным дисциплинам. Пособие включает в себя кратко изложенный теоретический материал с подробно разобранными на каждую тему примерами, что позволяет использовать его для организации и контроля самостоятельной работы студентов.

В 1 разделе рассматриваются понятия генеральной совокупности и выборки, методы их представления. Раздел 2 посвящен основным числовым характеристикам выборки и методам их вычисления. Нахождение числовых характеристик выборки, полученной в результате объединения нескольких выборок в одну, рассматривается в разделе 3. В 4 разделе разбираются интервальные оценки математического ожидания и дисперсии выборки. В разделе 5 рассматриваются понятия линейной корреляции, уравнения регрессии, а также ранговой корреляции. Большое внимание в работе посвящено проверке статистических гипотез (раздел 6). Во второй части работы (раздел 7) приведено 20 вариантов индивидуальных заданий.

Пособие может быть использовано для проведения практических и лабораторных занятий. При его написании были учтены современные требования и компетенции, предъявляемые к бакалаврам вышеуказанных направлений подготовки/специальностей.

1 Выборка. Графическое и табличное представление данных

Пусть из генеральной совокупности, обладающей количественным признаком X получена выборка n элементов $x_1,\ x_2,\ \ldots,\ x_n$. Расположив эти элементы в порядке возрастания, получим **вариаионный ряд**. Значения x_i называют **вариантами**. Обозначим x_{min}, x_{max} наименьшее и наибольшее значение элементов выборки. Величина

$$R = x_{max} - x_{min} \tag{1}$$

называется **размах вариации**. Для дальнейшей обработки данных целесообразно построить *интервальный вариационный ряд*, который задается в виде таблицы

$$\begin{array}{ccccc} x_i & [x_0; x_1) & [x_1; x_2) & \ldots & [x_{k-1}; x_k) \\ n_i & n_1 & n_2 & \ldots & n_k \end{array}. \tag{2}$$

Для определения количества интервалов k можно воспользоваться формулой Стерджеса $k = 1 + \log_2 n,$ которая носит не обязательный рекомендательный характер. Тогда длина частичных интервалов равна $h = \dfrac{R}{1 + \log_2 n}$. Обычно значение h округляют до ближайшего целого числа или ближайшей удобной для представления дроби. За начало первого интервала рекомендуется взять $x_{\text{нач}} = x_{min} - \dfrac{h}{2}$, а конец последнего интервала должен удовлетворять условию $x_{\text{кон}} - h \leq x_{max} < x_{\text{кон}}$. Промежуточные интервалы получают прибавляя величину h к концу предыдущего интервала. После этого подсчитывают количество вариант, попавших в каждый интервал $[x_{i-1}; x_i)$. Эти числа обозначим n_i и назовем **частотой** данного интервала, а $w_i = \dfrac{n_i}{n}$ — **относительной частотой**. Очевидно, что $\sum\limits_{i=1}^{k} n_i = n,\ \ \sum\limits_{i=1}^{k} w_i = 1.$

Пример 1. Проведено $n = 200$ измерений некоторой случайной величины.Оказалось что $x_{min} = 13,58,$ а $x_{max} = 13,73.$ Определить количество и границы интервалов при построении интервального вариационного ряда.

Решение. По формуле Стерджеса находим $h = \dfrac{13,73 - 13,58}{1 + \log_2 200} \approx 0,0174 \approx 0,02$. Тогда $x_{\text{нач}} = 13,58 - \dfrac{0,02}{2} = 13,57$ и получаем следующие интервалы: [13,57; 13,59), [13,59; 13,61), …, [13,73; 13,75). Далее, если даны результаты всех измерений, подсчитываем **частоты**, соответствующие каждому интервалу и строим **интервальный вариационный ряд**.

Замечание 1. В некоторых случаях соседние интервалы, в которых содержится мало вариант, объединяют в один так, чтобы соответствующая ему частота содержала не менее чем 5 – 8 вариант.

Замечание 2. Иногда по полученному интервальному ряду строят **дискретный вариационный ряд**, заменяя интервал $[x_{i-1}; x_i)$ на серединное значние этого интервала: $x_i^* = \dfrac{x_{i-1} + x_i}{2}$:

x_i^*	x_1^*	x_2^*	$\dots$	x_k^*
n_i	n_1	n_2	$\dots$	n_k

Для наглядности представления вариационных рядов строят **полигон частот (относительных частот)** для дискретного вариационного ряда или **гистограмму частот (относительных частот)** для интервального ряда.

Полигон частот это ломаная, отрезки которой соединяют точки $(x_1^*; n_1)$, $(x_2^*; n_2)$, …, $(x_k^*; n_k)$. Соответственно, график полигона относительных частот проходит через точки $(x_1^*; w_1)$, $(x_2^*; w_2)$, …, $(x_k^*; w_k)$.

Гистограмма частот (относительных частот) это ступенчатая фигура, состоящая из прямоугольников, основанием которых служат частичные интервалы длиной h, а высоты равны $\dfrac{n_i}{h}$ $\left(\dfrac{w_i}{h}\right)$.

График **накопленных частот** строится аналогично гистограмме с той лишь разницей, что для рассчета высот прямоугольников берутся накопленные частоты, равные $\sum\limits_{i=1}^{j} \dfrac{n_i}{h}$, $j = 1, 2, \dots, k$ или

$\sum\limits_{i=1}^{j} \dfrac{w_i}{h}$ для **накопленных относительных частот.**

График накопленных относительных частот близок по своему виду к графику **эмпирической функции распределения.** Для выборки объема n эмпирическая функция распределения $F^*(x)$ задается уравнением $F^*(x) = \dfrac{n_x}{n}$, где n_x – число вариант, меньших x.

2 Выборочные числовые характеристики и точечные оценки

Пусть дан дискретный вариационный ряд

$$\begin{array}{ccccc} x_i & x_1 & x_2 & \ldots & x_k \\ n_i & n_1 & n_2 & \ldots & n_k \end{array}. \tag{3}$$

.

Для дальнейшей обработки статистических данных применяются **выборочные числовые характеристики**, среди которых выделим следующие:

выборочное среднее

$\overline{x}_{\text{в}} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$ или $\overline{x}_{\text{в}} = \dfrac{1}{n}\sum\limits_{i=1}^{k} n_i\, x_i$, где $n = \sum\limits_{i=1}^{k} n_i$ — объем выборки;

выборочная дисперсия

$D_{\text{в}} = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x}_{\text{в}})^2$ или $D_{\text{в}} = \dfrac{1}{n}\sum\limits_{i=1}^{k} n_i\,(x_i - \overline{x}_{\text{в}})^2$.

Часто для нахождения дисперсии используют формулу

$$D_{\text{в}} = \overline{x^2} - \left(\overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\, x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{k} n_i\, x_i\right)^2. \tag{4}$$

.

среднее квадратическое отклонение $\sigma_{\text{в}} = \sqrt{D_{\text{в}}}$;

начальный выборочный момент k-го порядка

$M_k = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^k$;

центральный выборочный момент k-го порядка $m_k = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x}_{\text{в}})^k$.

Очевидно, что $\overline{x}_{\text{в}} = M_1$, $D_{\text{в}} = m_2$. Кроме перечисленных характеристик отметим следующие:

мода Mo— варианта, имеющая наибольшую частоту;

медиана Me — варианта, которая делит ряд на две равные по числу вариант части, если число вариант нечетное, или полусумме серединных вариант при четном числе элементов вариационного ряда.

коэффициент вариации $V = \dfrac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \cdot 100\%$;

асимметрия $a_s = \dfrac{m_3}{\sigma_{\text{в}}^3}$;

эксцесс $e_k = \dfrac{m_4}{\sigma_{\text{в}}^4} - 3$.

Мода и медиана *интервального* ряда с интервалами равной длины находятся по формулам:

$$Mo = x_{Mo} + \Delta \cdot \frac{m_2 - m_1}{(m_2 - m_1) + (m_2 - m_3)},$$

где x_{Mo} – начало модального интервала, Δ – длина интервала; m_1 – частота домодального, m_2 – модального, m_3 – замодального интервалов.

$$Me = x_{Me} + \Delta \cdot \frac{\frac{n}{2} - m_{\text{н}}}{m_{Me}},$$

где x_{Me} – начало медианного интервала; n – объем выборки; $m_{\text{н}}$ – накопленная частота интервала, предшествующего медианному; m_{Me} – локальная частота медианного интервала.

Пусть изучается генеральная совокупность, объекты которой обладают некоторым количественным признаком X. Проведя *сплошное исследование* можно получить полное представление о случайной величине X, построить закон распределения и найти его параметры. Но в силу ряда причин сделать это зачастую не представляется возможным. Поэтому на практике приходится иметь дело с *выборкой*, по которой и приходится судить о генеральной совокупности в целом.

Выборочные характеристики $\overline{x}_{\text{в}}$, $D_{\text{в}}$ и $\sigma_{\text{в}}$ являются **эффективными** и **состоятельными** точечными оценками соответствующих параметров генеральной совокупности. Причем, $\overline{x}_{\text{в}}$ — **несмещенная** оценка для генерального среднего $a = \overline{x}_{\text{в}}$. Несмещенной оценкой для генеральной дисперсии является **исправленная дисперсия** $s^2 = \dfrac{n}{n-1} \cdot D_{\text{в}} = \dfrac{1}{n-1} \sum\limits_{i=1}^{n} (x_i - \overline{x}_{\text{в}})^2$. В отличии от s^2 величина $s = \sqrt{s^2}$ остается смещенной оценкой генерального среднего квадратического отклонения, поэтому ее называют **"исправленное" среднее квадратическое отклонение**.

3 Объединение нескольких выборок в одну

Пусть выборка (или генеральная совокупность) объема n разбита на k групп, содержащих $n_1, n_2, \ldots n_k$ элементов соответственно, $\sum\limits_{j=1}^{k} n_j = n$. В каждой группе найдены средние $\overline{x}_j$ и дисперсии D_j. Тогда выборочное (или генеральное) среднее всей совокупности выражается через $\overline{x}_j$ равенством

$$\overline{x} = \frac{\sum\limits_{j=1}^{k} n_j \, \overline{x}_j}{n},$$

а общая дисперсия D равна сумме среднего арифметического групповых дисперсий $\overline{D}_{\text{гр}}$ и межгрупповой дисперсии $D_{\text{межгр}}$:

$$D = \overline{D}_{\text{гр}} + D_{\text{межгр}}, \tag{5}$$

где

$$\overline{D}_{\text{гр}} = \frac{\sum\limits_{j=1}^{k} n_j \, D_j}{n}; \qquad D_{\text{межгр}} = \frac{\sum\limits_{j=1}^{k} n_j \, (\overline{x}_j - \overline{x})^2}{n}.$$

Формулу (5) применяют в случае, если первичные данные не известны, а также при обработке большого массива данных, когда его разбивают на несколько групп, в каждой из которых вычисляют среднее значение и дисперсию.

Пример 2. По трём выборкам объёма $n_1 = 13$, $n_2 = 15$ и $n_3 = 7$ найдены выборочные средние $\overline{x}_{\text{в}1} = 4,20$, $\overline{x}_{\text{в}2} = 4,44$, $\overline{x}_{\text{в}3} = 4,12$. Эти выборки объединены в одну общую. Найти $\overline{x}_{\text{в}}$ объединенной выборки.

Решение.

$$\overline{x}_{\text{в}} = \frac{\sum\limits_{j=1}^{3} n_j \overline{x}_{\text{в}j}}{\sum\limits_{j=1}^{3} n_j} =$$

$$= \frac{n_1\overline{x}_{\text{в}1} + n_2\overline{x}_{\text{в}2} + n_3\overline{x}_{\text{в}3}}{n_1 + n_2 + n_3} = \frac{13 \cdot 4,20 + 15 \cdot 4,44 + 7 \cdot 4,12}{13 + 15 + 7} = 4,29$$

Пример 3. Пусть для выборок примера 2 найдены групповые дисперсии $D_{\text{в}1} = 1,32$; $D_{\text{в}2} = 1,04$; $D_{\text{в}3} = 1,37$. Выборки объединены в одну. Найти дисперсию объединённой выборки.

Решение. Найдём среднюю групповую дисперсию

$$\overline{D}_{\text{гр}} = \frac{\sum\limits_{j=1}^{k} n_j\, D_{\text{в}j}}{\sum n_j} = \frac{13 \cdot 1,32 + 15 \cdot 1,04 + 7 \cdot 1,37}{35} = 1,21.$$

Найдём межгрупповую дисперсию

$$D_{\text{межгр}} = \frac{\sum\limits_{j=1}^{k} n_j\, (\overline{x}_{\text{в}j} - \overline{x})^2}{\sum n_j} =$$

$$= \frac{13 \cdot (4,2 - 4,29)^2 + 15 \cdot (4,44 - 4,29)^2 + 7 \cdot (4,12 - 4,29)^2}{35} = 0.02.$$

Тогда $D = \overline{D}_{\text{гр}} + D_{\text{межгр.}} = 1.21 + 0,02 = 1,23.$

Заметим, что при вычислении дисперсии можно было воспользоваться формулой (4).

Ответ: Дисперсия объединенной выборки равна $D = 1,23$.

4 Интервальные оценки

Состоятельные точечные оценки с увеличением объема выборки n хотя и стремятся по вероятности к точному значению оцениваемого параметра, но тем не менее остаются приближенными значениями. Поэтому естественно возникает вопрос о точности полученой оценки. В математической статистике этот вопрос формулируется следующим образом: найти интервал, который с заданной вероятностью $0 < \gamma < 1$ содержит внутри себя точное значение оцениваемого параметра. То есть если θ^* – точечная оценка некоторого параметра θ и выполнено условие $P(|\theta^* - \theta| < \delta) = \gamma$, то интервыл $\theta^* - \delta < \theta < \theta^* + \delta$ называется **доверительный**, а вероятность γ — **надежностью оценки.** Чем при заданной надежности меньше δ, тем выше **точность оценки** θ^* параметра θ.

4.1 Интервальные оценки математического ожидания

1. Для оценки **математического ожидания (генеральной средней)** $a = \overline{x}_\text{г}$ нормально распределенного количественного признака X по выборочной средней $\overline{x}_\text{в}$ при известном среднем квадратическом отклонении генеральной совокупности σ при повторном отборе доверительный интервал, который с надежностью γ содержит $\overline{x}_\text{в}$, задается формулой

$$\overline{x}_\text{в} - \frac{t \cdot \sigma}{\sqrt{n}} < a < \overline{x}_\text{в} + \frac{t \cdot \sigma}{\sqrt{n}}, \tag{6}$$

где t находится из равенства

$$2\Phi(t) = \gamma. \tag{7}$$

Здесь $\Phi(t) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$ – функция Лапласа.

Запишем формулу (6) в виде $\left|\overline{x}_\text{в} - a\right| < \delta = \dfrac{t \cdot \sigma}{\sqrt{n}}$. Тогда $\overline{x}_\text{в}$ – точечная оценка математического ожидания a генеральной совокупности, а величина $\delta = \dfrac{t \cdot \sigma}{\sqrt{n}}$ – точность этой оценки. Отсюда можно получить следущий результат.

Минимальный объем выборки, при котором с заданной надежностью γ точность оценки не превосходит δ, удовлетворяет неравенству

$$n \geq \frac{t^2 \cdot \sigma^2}{\delta^2}, \tag{8}$$

где t находится из условия (7).

2. Если выборка **бесповторная** и N — объем гененеральной совокупности, то формула (6) примет вид

$$\overline{x}_{\text{в}} - \frac{t \cdot \sigma}{\sqrt{n}}\sqrt{1 - \frac{n}{N}} < a < \overline{x}_{\text{в}} + \frac{t \cdot \sigma}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}, \tag{9}$$

где $\delta = \dfrac{t \cdot \sigma}{\sqrt{n}}\sqrt{1 - \dfrac{n}{N}}$. Заметим, что при увеличении N множитель $\sqrt{1 - \dfrac{n}{N}}$ стремится к 1, а отличие между повторной и бесповторной выборками сглаживается. Поэтому при больших N и для бесповторных выборок пользуются формулой (6).

3. Пусть как и прежде количественный признак X генеральной совокупности распределен нормально, но σ не известно. По выборке объема n найдено $\overline{x}_{\text{в}}$, и "исправленное" среднее квадратическое отклонение s. Тогда доверительный интервал для оценки математического ожидания a при повторном отборе задается формулой

$$\overline{x}_{\text{в}} - \frac{t_\gamma \cdot s}{\sqrt{n}} < a < \overline{x}_{\text{в}} + \frac{t_\gamma \cdot s}{\sqrt{n}}, \tag{10}$$

где t_γ определяется по таблицам распределения Стьюдента по уровню значимости $\alpha = 1 - \gamma$ для двусторонней критической области и числу степеней свободы $k = n-1$; s — "исправленное" среднее квадратическое отклонение, n — объем выборки, $\delta = \dfrac{ts}{\sqrt{n}}$.

Если объем выборки N и выборка бесповторная, то доверительный интервал задается формулой

$$\overline{x}_{\text{в}} - \frac{t_\gamma \cdot s}{\sqrt{n}}\sqrt{1-\frac{n}{N}} < a < \overline{x}_{\text{в}} + \frac{t_\gamma \cdot s}{\sqrt{n}}\sqrt{1-\frac{n}{N}}. \qquad (11)$$

Здесь $\delta = \dfrac{t_\gamma \cdot s}{\sqrt{n}}\sqrt{1-\dfrac{n}{N}}$.

Так как при увеличении объема выборки n распределение Стьюдента медленно приближается к нормальному, то при $n > 30$ в формулах (10) и (11) вместо t_γ поставить t, найденное из равенства $2\Phi(t) = \gamma$ и пренебречь множителем $\sqrt{1-\dfrac{n}{N}}$.

4.2 Интервальные оценки дисперсии и среднего квадратического отклонения

Пусть s^2 — исправленная дисперсия, а s — “исправленное “ среднее квадратическое отклонение выборки. Тогда для генеральных дисперсии и среднего квадратического отклонения справедливы оценки

$$\frac{(n-1)\cdot s^2}{\chi^2_{\frac{1-\gamma}{2};\,n-1}} < \sigma^2 < \frac{(n-1)\cdot s^2}{\chi^2_{\frac{1+\gamma}{2};\,n-1}}$$

и

$$s\cdot\sqrt{\frac{n-1}{\chi^2_{\frac{1-\gamma}{2};\,n-1}}} < \sigma < s\cdot\sqrt{\frac{n-1}{\chi^2_{\frac{1+\gamma}{2};\,n-1}}}, \qquad (12)$$

где $\chi^2_{\frac{\alpha}{2},n-1}$ и $\chi^2_{1-\frac{\alpha}{2},n-1}$ — критические точки распределения χ^2 с $n-1$ стапенями свободы и соответствующими уровнями значимости при $\alpha = 1-\gamma$.

Пример 4. Из генеральной совокупности большого объема, подчиненной нормальному закону распределения взята выборка объема $n = 20$. По этой выборке найдены выборочная средняя $\overline{x}_{\text{в}} = 45$ и выборочная дисперсия. $D_{\text{в}} = 4$. Найти доверительные интервалы для a) оценки с надежностью $\gamma = 0,95$ генеральной средней $\overline{x}_{\text{г}}$; b) генерального среднего квадратического отклонения σ.

Решение. a) Так как σ не известна, оценку генеральной средней получим по формуле (10). Найдем "исправленное" среднее квадратическое отклонение

$$s = \sqrt{s^2} = \sqrt{\frac{n}{n-1} \cdot D_{\text{в}}} = \sqrt{\frac{20}{19} \cdot 4} = 2,05.$$

По таблице значений t_γ найдем $t_\gamma = t_{0,95;\,20} = 2,093$. Тогда в силу формулы (10) получим

$$45 - \frac{2,093 \cdot 2,05}{\sqrt{20}} < \overline{x}_{\text{г}} < 45 + \frac{2,093 \cdot 2,05}{\sqrt{20}}, \quad \text{или} \quad 44,04 < \overline{x}_{\text{г}} < 45,96.$$

b) Найдем доверительный интервал для σ. По таблице критических точек распределения χ^2 получим $\chi^2_{\frac{1-0,95}{2};\,20-1} = 32,9$; $\chi^2_{\frac{1+0,95}{2};\,20-1} = 8,91$. Применим формулу (12).

$$2,05 \cdot \sqrt{\frac{20-1}{32,9}} < \sigma < 2,05 \cdot \sqrt{\frac{20-1}{8,91}}, \quad \text{или} \quad 1,56 < \sigma < 2,99.$$

Ответ: С надежностью $\gamma = 0,95$ $44,04 < \overline{x}_{\text{г}} < 45,96$, $1,56 < \sigma < 2,99$.

Пример 5. Найти минимальный объем выборки, при котором с надежностью $\gamma = 0,901$ точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta = 2$, если известно среднее квадратическое отклонение $\sigma = 9$ нормально распределенной генеральной совокупности.

Решение. Из равенства $2\Phi(t) = 0,901$ по таблице функции Лапласа находим $t = 1,65$. Тогда, в силу формулы (8), получим

$$n \geq \frac{1,65^2 \cdot 9^2}{2^2} = 55,13$$

Округлив результат по избытку, получим $n = 56$.

Ответ: Надо взять выборку объема $n = 56$.

5 Корреляционная зависимость

А. Коэффициент корреляции Пирсона. Пусть даны две случайные случайные величины X и Y. Зависимость, при которой закон распределения одной из них зависит от того, какие значения приняла другая величина, называется **статистической зависимостью.** Если при этом изменение одной величины влечет за собой изменение среднего значения другой, то зависимость называется **корреляционной.** В этом случае можно говорить о функции, которая каждому значению x ставит в соответствие среднее значение $\overline{y}_x$: $\overline{y}_x = f^*(x)$. Полученное уравнение называют **уравнением регрессии** Y на X. Для построения уравнения регрессии применяется **метод наименьших квадратов.** Пусть выборочные данные представлены в виде корреляционной таблицы:

			X		
Y	x_1	x_2	$\cdots$	x_p	n_y
y_1	$n_{x_1y_1}$	$n_{x_2y_1}$	$\cdots$	$n_{x_py_1}$	n_{y_1}
y_2	$n_{x_1y_2}$	$n_{x_2y_2}$	$\cdots$	$n_{x_py_2}$	n_{y_2}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_m	$n_{x_1y_m}$	$n_{x_2y_m}$	$\cdots$	$n_{x_py_m}$	n_{y_m}
n_x	n_{x_1}	n_{x_2}	$\cdots$	n_{x_p}	n

здесь $n_{x_iy_j}$ — частота, с которой встречается пара наблюдений x_i, y_j; $n_{x_i} = \sum\limits_{j=1}^{m} n_{x_iy_j}$; $n_{y_j} = \sum\limits_{i=1}^{p} n_{x_iy_j}$; $n = \sum\limits_{i=1}^{p} n_{x_i} = \sum\limits_{j=1}^{m} n_{y_j}$ — объем выборки. Рассмотрим частный случай, когда уравнение регрессии линейно: $\overline{y}_x = \rho_{yx}x + b$. Тогда коэффициенты ρ_{yx} и b находятся из системы линейных уравнений

$$\begin{cases} (n\overline{x^2}) \cdot \rho_{yx} & + & (n\overline{x}) \cdot b & = & \sum\limits_{i=1,j=1}^{p,m} (n_{x_iy_j} \cdot x_i \cdot y_j), \\ (\overline{x}) \cdot \rho_{yx} & + & b & = & y. \end{cases}$$

Уравнение линейной регрессии можно искать в виде

$$\overline{y}_x - \overline{y} = r_{\text{в}} \frac{\sigma_y}{\sigma_x}(x - \overline{x}), \tag{13}$$

где

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_{x_i} x_i, \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{m} n_{y_j} y_j, \quad \text{— выборочные средние,} \tag{14}$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{p} n_{x_i} x_i^2, \ \overline{y^2} = \frac{1}{n}\sum_{j=1}^{m} n_{y_j} y_j^2, \ \sigma_x = \sqrt{\overline{x^2} - \overline{x}^2}, \ \sigma_y = \sqrt{\overline{y^2} - \overline{y}^2}, \tag{15}$$

$$K_{x;\,y} = \frac{\sum\limits_{i=1,j=1}^{p,\,m} (n_{x_i y_j} \cdot x_i \cdot y_j)}{n} - \overline{x} \cdot \overline{y} \quad \text{– коэффициент ковариации,} \tag{16}$$

$$r_{\text{в}} = \frac{\sum\limits_{i=1,j=1}^{p,\,m} (n_{x_i y_j} \cdot x_i \cdot y_j) - n \cdot \overline{x} \cdot \overline{y}}{n \cdot \sigma_x \cdot \sigma_y} \quad \text{или} \quad r_{\text{в}} = \frac{K_{x;\,y}}{\sigma_x \cdot \sigma_y}. \tag{17}$$

Коэффициент $r_{\text{в}}$ является точечной оценкой генерального значения коэффициента корреляции Пирсона r, который является важной характеристикой двумерной случайной величины. Он характеризует тесноту связи между X и Y. Можно показать, что $|r| \le 1$ и чем теснее связь между признаками X и Y, тем ближе к 1 величина $|r|$. Если Y и X связаны линейной функциональной зависимостью, то $r = \pm 1$. Если же Y и X независимы, то $r = 0$.

Пример 6. По выборке, извлеченной из двумерной нормальной генеральной совокупности (X, Y), составлена корреляционная таблица. Найти выборочный коэффициент корреляции и написать уравнение регрессии Y на X.

			X			
Y	$0,5$	$0,7$	$0,9$	$1,1$	$1,3$	n_y
4	–	–	2	21	2	25
7	2	4	12	14	–	32
10	–	2	3	–	–	5
13	8	9	1	–	–	18
n_x	10	15	18	35	2	$n=80$

Решение. Найдем параметры уравнения регрессии по формулам (14) – (17):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_{x_i}x_i = \\ = \frac{0,5\cdot 10 + 0,7\cdot 15 + 0,9\cdot 18 + 1,1\cdot 35 + 1,3\cdot 2}{80} = \frac{72,8}{80} = 0,91,$$

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{m} n_{y_j}y_j = \frac{4\cdot 25 + 7\cdot 32 + 10\cdot 5 + 13\cdot 18}{80} = \frac{608}{80} = 7,6,$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{p} n_{x_i}x_i^2 = \\ = \frac{0,5^2\cdot 10 + 0,7^2\cdot 15 + 0,9^2\cdot 18 + 1,1^2\cdot 35 + 1,3^2\cdot 2}{80} = 0,877$$

$$\overline{y^2} = \frac{1}{n}\sum_{j=1}^{m} n_{y_j}y_j^2 = \frac{4^2\cdot 25 + 7^2\cdot 32 + 10^2\cdot 5 + 13^2\cdot 18}{80} = 68,875$$

$$\sigma_x = \sqrt{\overline{x^2} - \overline{x}^2} = \sqrt{0,877 - 0,91^2} = 0,2211 \approx 0,22$$

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{68,875 - 7,6^2} = 3,3339 \approx 3,33$$

$$\frac{\sum\limits_{i=1,j=1}^{p,m}(n_{x_iy_j}\cdot x_i\cdot y_j)}{n} = \frac{1}{80}(0,9\cdot 4\cdot 2 + 1,1\cdot 4\cdot 21 + 1,3\cdot 4\cdot 2 + \\ +0,5\cdot 7\cdot 2 + 0,7\cdot 7\cdot 4 + 0,9\cdot 7\cdot 12 + 1,1\cdot 7\cdot 14 + 0,7\cdot 10\cdot 2 + \\ +0,9\cdot 10\cdot 3 + 0,5\cdot 13\cdot 8 + 0,7\cdot 13\cdot 9 + 0,9\cdot 13\cdot 1) = 6,3325$$

$$K_{x;\,y}=\frac{\sum\limits_{i=1,j=1}^{p,\,m}(n_{x_iy_j}\cdot x_i\cdot y_j)}{n}-\overline{x}\cdot\overline{y}=6,3325-0,91\cdot 7,6=-0,5835$$

$$r_{\text{в}}=\frac{K_{x;\,y}}{\sigma_x\cdot\sigma_y}=\frac{-0,5835}{0,22\cdot 3,33}=-0,796\approx -0,80$$

Подставив найденные значения в формулу (13), получим уравнение регрессии

$$\overline{y}_x-7,6=-0,80\cdot\frac{3,33}{0,22}(x-0,91),\quad \text{или}\quad \overline{y}_x=-12,1x+18,62.$$

Ответ: Коэффициент корреляции $r_{\text{в}}=-0,8$; уравнение регрессии $\overline{y}_x=-12,1x+18,62$.

В. Коэффициент ранговой корреляции Спирмена. Из генеральной совокупности, каждый объект которой обладает двумя различимыми, не важно, количественными или качественными признаками X и Y, получена выборка объема n. Расположим объекты один за другим в порядке изменения выраженности признака X и прсвоим им порядковые номера **(ранги).** Объектам, не различимым по степени выраженности признака, присваиваются одинаковые ранги, равные среднему арифметическому их порядковых номеров. Далее, аналогичным образом определим ранги по признаку Y. Таким образом каждому объекту ставится в соотвествие пара чисел — их ранги $(x_1,\,y_i)$. Коэффициенткорреляции Пирсона, вычисленный для полученной двумерной случайной величины называется **коэффициентом ранговой корреляции Спирмена:**

$$\rho_{\text{в}}=1-\frac{6\sum\limits_{i=1}^{n}(x_i-y_i)^2}{n^3-n}. \qquad (18)$$

Этот коэффициент, так же, как и коэффициент Пирсона, отражает величину тесноты связи между признаками X и Y.

Пример 7. Денадцать студентов написали контрольную работу, после чего прошли тестирование на компьютере. Результаты контрольной по 30-балльной шкале даны в первой строке, тестирования (максимум 100 баллов) во второй строке. Найти коэффициент ранговой корреляции Спирмена между результатами кнтрольной и тестирования.

контр	30	23	27	24	18	28	28	20	25	23	18	23
тест	96	71	87	80	64	92	95	60	87	88	62	72

Решение. Переставим столбцы так, чтобы баллы за контрольную располагались в порядке возрастания и присвоим ранги x_i (запишем в верхнюю строку) согласно порядковым номерам столбиков. Аналогично проранжируем баллы по тесту. Получим ранги y_i (нижняя строка).

ранг x_i	1	$2,5$	$2,5$	4	5	6	8	8	8	10	$11,5$	$11,5$
контр	30	(28	28)	27	25	24	(23	23	23)	20	(18	18)
тест	96	92	95	87	87	80	88	72	71	60	62	64
ранг y_i	1	3	2	$5,5$	$5,5$	7	4	8	9	12	11	10

Найдем сумму квадратов разностей рангов:

ранг x_i	1	$2,5$	$2,5$	4	5	6	8	8	8	10	$11,5$	$11,5$
ранг y_i	1	3	2	$5,5$	$5,5$	7	4	8	9	12	11	10
$(x_i - y_i)^2$	0	$0,5^2$	$0,5^2$	$1,5^2$	$1,5^2$	1	4^2	0	1	2^2	$0,5^2$	$0,5^2$

Найдем сумму элементов последней строки: $\sum(x_i - y_i)^2 = 27,5$. Подставив значения в формулу (18), получим

$$\rho_{\text{в}} = 1 - \frac{6\sum\limits_{i=1}^{n}(x_i - y_i)^2}{n^3 - n} = 1 - \frac{6 \cdot 27,5}{12^3 - 12} = 0,904.$$

Ответ: Коэффициент ранговой корреляции $\rho_{\text{в}} = 0,904$

6 Статистическая проверка статистических гипотез

Статистической гипотезой называется предположение о величине параметров известного распределения или о законе неизвестного распределения.

Рассмотрим понятие статистической гипотезы и ее статистической проверки на примере следущих рассуждений. Пусть из двух генеральных совокупностей взяты выборки, по которым найдены значения некоторого параметра, например, исправленная дисперсия s_1^2 и s_2^2. Пусть есть основания полагать, что дисперсии в генеральных совокупностях равны между собой, то есть $D_1 = D_2$. Сформулируем это предположение в виде основной гипотезы $H_0 : \quad D_1 = D_2$. Отрицание этого утверждения назовем альтернативной гипотезой и обозначим $H_1 : \quad D_1 \neq D_2$. Для проверки гипотезы H_0 найдем отношение $F_{\text{набл}} = \dfrac{s_1^2}{s_2^2}$, величина которого зависит от конкретных данных, попавших в выборку. Поэтому $F_{\text{набл}}$ есть наблюдаемое значение случайной величины $F = \dfrac{S_1^2}{S_2^2}$ отношения исправленных дисперсий. Очевидно, чем сильнее $F_{\text{набл}}$ отличается от 1, тем меньше вероятность того, что гипотеза H_0 верна. Величину F назовем критерием и можно говорить о законе распределения этой случайной величины. И если этот закон построить, то можно рассмотреть следущую задачу:

Пусть H_0 справедливо. Найти $F_{\text{кр}}$ такое, чтобы для заданного достаточно малого α выполнялось неравенство $P(F > F\text{кр}) < \alpha$. Тогда если $F_{\text{набл}} > F_{\text{кр}}$, то это будет означать, что произошло маловероятное событие, то есть скорее всего гипотеза H_0 не верна. В этом случае гипотезу H_0 отвергают и принимают альтернативную гипотезу H_1. При этом величину вероятности α называют уровнем значимости критерия F. Если $F_{\text{набл}} < F_{\text{кр}}$, гипотезу H_0 принимают, но ответ формулируют более осторожно: При заданном уровне значимости α нет оснований отвергать гипотезу H_0.

Другой пример. Пусть n_i – эмпирические частоты некоторого распределения, полученные в результате наблюдений. Пусть есть основания полагать, что генеральная совокупность подчиняется закону (например, нормальному), по которому найдены теоретические частоты n'_i, и $\sum n_i = \sum n'_i$.

$$\begin{array}{lllll} x_i & x_1 & x_2 & \dots & x_k \\ n_i & n_1 & n_2 & \dots & n_k. \\ n'_i & n'_1 & n'_2 & \dots & n'_k \end{array}$$

Основная гипотеза $H_0:$ генеральная совокупность подчиняется нормальному закону. Конкурирующая гипотеза H_1 : распределение нормальным не является. Для проверки гипотезы построим критерий $\chi^2 = \sum\limits_{i=1}^{k} \dfrac{(n_i - n'_i)^2}{n'_i}$. Чем сильнее отличаются эмпирические частоты от теоретических, тем больше значение критерия χ^2 и тем меньше вероятность справедливости гипотезы H_0. Зная закон расределения χ^2, можно для заданной вероятности α найти $\chi^2_{\text{кр}}$ такое, что при $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$ вероятность справедливости гипотезы H_0 будет меньше α. В этом случае гипотезу H_0 при заданном уровне значимости α отвергают. Если же $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, как и в прошлом примере говорят, что при заданном уровне значимости α нет оснований отвергать гипотезу H_0.

6.1 Гипотеза о равенстве дисперсий двух нормальных генеральных совокупностей

Пусть по независимым выборкам объема n_1 и n_2 из нормальных генеральных совокупностей найдены исправленные выборочные дисперсии s^2_X и s^2_Y. Выдвинем гипотезу $H_0: \quad D_X = D_Y$ (дисперсии генеральных совокупностей равны между собой). Альтернативная гипотеза H_1 может быть сформулирована в виде $H_1: \; D_X \neq D_Y$, $D_X < D_Y$ или $D_X > D_Y$.

Для проверки гипотезы H_0 строится критерий

$$F = \frac{S^2_{\text{б}}}{S^2_{\text{м}}}.$$

Эта случайная величина подчиняется закону распределения Фишера – Снедекора с $k_1 = n_б - 1$, $k_2 = n_м - 1$ степенями свободы. Здесь $S_б = \max(S_1, S_2)$, $S_{\,м} = \min(S_1, S_2)$, $n_б$ и $n_м$ — объемы выборок, по которым найдены $S_б$ и $S_м$ соответственно.

а) Для проверки $H_0 : D_X = D_Y \;\; H_1 : D_X > D_Y$ при заданном уровне значимости α надо вычислить

$$F_{набл} = \frac{s_б^2}{s_м^2}$$

и по таблице критических точек распределения функции Фишера – Снедекора, по заданному уровню значимости α и числам степеней свободы k_1 и k_2 найти критическую точку $F_{кр} = F(\alpha, k_1, k_2)$. Если $F_{набл} < F_{кр}$, нет оснований отвергать гипотезу H_0; при $F_{набл} > F_{кр}$ гипотезу H_0 отвергают.

б) Если $H_0 : D_X = D_Y \;\; H_1 : D_X \neq D_Y$ то критическую точку $F_{кр}$ находят из условия $F_{кр} = F_{кр}(\frac{\alpha}{2}, k_1, k_2)$.

Пример 8. По двум независимым выборкам, объемы которых $n_X = 15$ и $n_Y = 12$ извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $s_X^2 = 43$ и $s_Y^2 = 91$. При уровне значимости $\alpha = 0,05$ проверить гипотезу $H_0 : D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1 : D(X) < D(Y)$.

Решение. У нас $s_Y^2 = 91 > 43 = s_X^2$, то есть $s_Y^2 > s_X^2$, то $s_б^2 = s_Y^2 = 91$, $s_м^2 = s_X^2 = 43$. Тогда $F_{набл} = \dfrac{s_б^2}{s_м^2} = \dfrac{91}{43} = 2,12$ Конкурирующая гипотеза $D_Y > D_X$, поэтому имеем правостороннюю критическую область. По таблице критических точек распределения F Фишера - Снедекора по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = n_б - 1 = n_Y - 1 = 12 - 1 = 11$ и $k_2 = n_м - 1 = n_X - 1 = 15 - 1 = 14$ найдем критическую точку $F_{кр}(0,05;\, 11;\, 14) = 2,56$. Так как $F_{набл} < F_{кр}$, нет оснований отвергать гипотезу H_0.

Ответ: Выборочные исправленные дисперсии различаются не значимо.

6.2 Сравнение наблюдаемой относительной частоты с гипотетической вероятностью появления события

По большому числу n независимых испытаний, проведенных по схеме Бернулли найдена относительная частота $\frac{m}{n}$. Вероятность p появления события A в генеральной совокупности неизвестна, но есть основания полагать, что она равна числу p_0. Требуется при заданном уровне значимости α проверить гипотезу $H_0 : p = p_0$ при конкурирующей гипотезе $H_1 : p \neq p_0$, $(H_1 : p > p_0$ или $H_1 : p < p_0)$.

Для проверки гипотезы H_0 строится критерий

$$U = \frac{\left(\dfrac{M}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0(1-p_0)}},$$

где M — случайная величина числа благоприятных исходов в выборке. При справедливости гипотезы H_0 при больших значениях n, то есть если $n > \dfrac{9}{p_0(1-p_0)}$, величина U имеет распределение близкое к нормальному $N(0; 1)$. Тогда:

a) Для проверки $H_0 : p = p_0$, $H_1 : p \neq p_0$ при заданном уровне значимости α надо вычислить

$$U_{\text{набл}} = \frac{\dfrac{m}{n} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \tag{19}$$

и по таблице функции $\Phi(x)$ найти $u_{\text{кр}}$ из равенства $\Phi(u_{\text{кр}}) = \dfrac{1-\alpha}{2}$. Если $|U_{\text{набл}}| < u_{\text{кр}}$, нет оснований отвергать гипотезу H_0; при $|U_{\text{набл}}| > u_{\text{кр}}$ гипотезу H_0 отвергают.

b) Если $H_0 : p = p_0$, $H_1 : p > p_0$, то критическую точку $u_{\text{кр}}^{\text{прав}}$ находят из условия $\Phi(u_{\text{кр}}^{\text{прав}}) = \dfrac{1-2\alpha}{2}$. При $U_{\text{набл}} > u_{\text{кр}}^{\text{прав}}$ гипотезу H_0 отвергают. При $U_{\text{набл}} < u_{\text{кр}}^{\text{прав}}$ нет оснований отвергать гипотезу H_0.

c) Если $H_0 : p = p_0$, $H_1 : p < p_0$, то находят $u_{\text{кр}}^{\text{лев}} = -u_{\text{кр}}^{\text{прав}}$. При $U_{\text{набл}} < u_{\text{кр}}^{\text{лев}}$ гипотезу H_0 отвергают.

Пример 9. Партия товара принимается к реализации, если количество бракованных изделий в ней не превосходит $P = 4\%$. По выборке объема $n = 250$ брак составил $m = 16$ изделий. Проверить при уровне значимости $\alpha = 0,015$, следует ли принять эту партию к реализации?

Решение. Заметим, что $np_0(1-p_0) = 250 \cdot 0,04 \cdot 0,96 = 9,6 > 9$, то есть объем выборки достаточно дольшой и мы можем воспользоваться рассмотренным выше критерием. Построим основную гипотезу: принять партию товара к реализации при альтернативной гипотезе: не принимать.

Партия будет принята, если $H_0 : \ p = p_0 = 0,04$ и отклонена, если $H_1 \ : \ p > p_0$. Имеем правостороннюю критическую область. Тогда по формуле (19) находим

$$U_{\text{набл}} = \frac{\dfrac{16}{250} - 0,04}{\sqrt{\dfrac{0,04 \cdot 0,96}{250}}} = 1,94.$$

Найдем $u_{\text{кр}}$ из равенства $\Phi(u_{\text{кр}}) = \dfrac{1-2\alpha}{2} = \dfrac{1-0,03}{2} = 0,485$. Тогда по таблице значений функции $\Phi(x)$ получим $u_{\text{кр}} = 2,17$. Так как $U_{\text{набл}} = 1,94 < 2,17 = u_{\text{кр}}$, нет оснований отвергать гипотезу H_0.

Ответ: Партию товара следует при уравне значимости $\alpha = 0,015$ принять к реализации.

6.3 Гипотеза о равенстве долей признака в двух совокупностях

Числовая характкрисстика, показывающая, какая часть генеральной совокупности обладает некоторым (качественным) признаком X называется долей этого признака. Эта характеристика является аналогом вероятности появления признака в единичном испытании. Точечной оценкой доли является величина $w = \dfrac{m}{n}$.

Пусть имеются две генеральные совокупности, пусть вероятность появления признака A в первой генеральной совокупности равна p_1, а во второй – p_2. Эти вероятности реально существуют, но нам неизвестны. По независимым выборкам большого объема n_1 и n_2 $(n_{1,2} \geq 100)$ из этих генеральных совокупностей найдены выборочные относительные частоты $w_1 = \dfrac{m_1}{n_1}$ и $w_2 = \dfrac{m_2}{n_2}$, где m_1 и m_2 – соответственно число элементов первой и второй выборок, обладающих признаком A. При уровне значимости α проверяется гипотеза $H_0 : p_1 = p_2$ (доли признака A в генеральных совокупностях равны) при конкурирующей гипотезе $H_1 : p_1 \neq p_2$ ($H_1 : p_1 > p_2$ или $H_1 : p_1 < p_2$.)

Для проверки гипотезы H_0 строится критерий

$$U = \frac{\dfrac{M_1}{n_1} - \dfrac{M_2}{n_2}}{\sqrt{p(1-p)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}},$$

где $p = p_1 = p_2$. При справедливости гипотезы H_0 Величина U имеет распределение близкое к нормальному $N(0;1)$. Так как p неизвестно, то вместо p можно взять ее оценку $p^* = \dfrac{m_1 + m_2}{n_1 + n_2}$. В результате получим:

а) Для проверки $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$ при заданном уровне значимости α надо вычислить

$$U_{\text{набл}} = \frac{\dfrac{m_1}{n_1} - \dfrac{m_2}{n_2}}{\sqrt{\dfrac{m_1 + m_2}{n_1 + n_2}\left(1 - \dfrac{m_1 + m_2}{n_1 + n_2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \tag{20}$$

и по таблице функции $\Phi(x)$ найти $u_{\text{кр}}$ из равенства $\Phi(u_{\text{кр}}) = \dfrac{1-\alpha}{2}$. Если $|U_{\text{набл}}| < u_{\text{кр}}$, нет оснований отвергать гипотезу H_0; при $|U_{\text{набл}}| > u_{\text{кр}}$ гипотезу H_0 отвергают.

б) Если $H_0 : p_1 = p_2$, $H_1 : p_1 > p_2$ или $H_1 : p_1 < p_2$, то критическую точку $u_{\text{кр}}$ находят из условия $\Phi(u_{\text{кр}}) = \dfrac{1-2\alpha}{2}$.

6.4 Сравнение долей признака в нескольких совокупностях

Пусть имеется r генеральных совокупностей, пусть вероятность появления признака A в каждой из них равна p_i, $i = 1, ..., r$. Эти вероятности реально существуют, но нам неизвестны. По r независимым выборкам большого объема $n_1, n_2 ..., n_r$ из этих генеральных совокупностей найдены выборочные относительные частоты $w_1 = \dfrac{m_1}{n_1}$, $w_2 = \dfrac{m_2}{n_2}, ..., w_r = \frac{m_r}{n_r}$, где m_1, m_2, ..., m_r – соответственно число элементов этих выборок, обладающих признаком A. При уровне значимости α проверяется гипотеза $H_0 : p_1 = p_2 = ... = p_r$ (обозначим это общее значение вероятности p) при конкурирующей гипотезе H_1 : равенство гипотезы H_0 нарушается (хотя бы в одном месте).

Для проверки гипотезы H_0 строится критерий

$$\chi^2 = \frac{\sum\limits_{i=1}^{r} n_i(W_i - p)^2}{p(1-p)},$$

где $W_i = \dfrac{M_i}{n_i}$, M_i — случайные величины, число благонриятных исходов в $i^{\text{й}}$ выборке. При справедливости гипотезы H_0 и при $n \to \infty$ Величина χ^2 имеет распределение Пирсона (распределение χ^2) с $k = r - 1$ степенями свободы.

В качестве оценки неизвестного значения p берут величину

$p_0 = \dfrac{\sum\limits_{i=1}^{r} m_i}{\sum\limits_{i=1}^{r} n_i}$ и вычисляют наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \frac{\sum\limits_{i=1}^{r} n_i(w_i - p_0)^2}{p_0(1-p_0)}. \qquad (21)$$

Для проверки гипотезы H_0 при конкурирующей гипотезе H_1 обычно берут правостороннюю критическую область. Гипотеза H_0 при заданном уровне значимости α отвергается, если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}} = \chi^2(\alpha; k)$, где $\chi^2_{\text{кр}}$ находится по таблице критических точек распределения χ^2 по уровню значимости α и числу степеней свободы $k = r - 1$.

Пример 10. Проведен опрос среди избирателей четырех возрастных групп. Из n_i опрошенных в $i^{\text{й}}$ группе пойти на выборы планируют m_i человек. Пусть $n_1 = 103$, $m_1 = 62$; $n_2 = 127$, $m_3 = 69$; $n_3 = 129$, $m_3 = 69$; $n_4 = 139$, $m_4 = 95$. При уровне значимости $\alpha = 0,05$ проверить гипотезы: a) доли желающих принять участие в выборах в первой и второй возрастной группе равны, альтернативная гипотеза: во второй группе эта доля выше; b) предполагаемая активность на выборах всех четырех групп одинакова.

Решение. **a)** $H_0 : p_1 = p_2$, $H_1 : p_1 > p_2$. Найдем $U_{\text{набл}}$ по формуле (20): $w_1 = \dfrac{m_1}{n_1} = \dfrac{62}{103} = 0,602$; $w_2 = \dfrac{m_2}{n_2} = \dfrac{67}{127} = 0,528$; $p = \dfrac{m_1 + m_2}{n_1 + n_2} = \dfrac{62 + 67}{103 + 127} = 0,561$;

$$U_{\text{набл}} = \frac{0,602 - 0,528}{\sqrt{0,561(1-0,561)\left(\dfrac{1}{103} + \dfrac{1}{127}\right)}} = 1,12.$$

Найдем $u_{\text{кр}}$ из условия $\Phi(u_{\text{кр}}) = \dfrac{1-2\alpha}{2} = \dfrac{1 - 2\cdot 0,05}{2} = 0,45$. По таблице функции $\Phi(x)$ находим $u_{\text{кр}} = 1,645$.

Так как $U_{\text{набл}} = 1,12 < 1,645 = u_{\text{кр}}$, нет оснований отвергать гипотезу H_0.

b) $H_0 : p_1 = p_2 = p_3 = p_4$, H_1 : равенство гипотезы H_0 нарушается (хотя бы в одном месте).

$$p_0 = \frac{\sum\limits_{i=1}^{4} m_i}{\sum\limits_{i=1}^{4} n_i} = \frac{62 + 67 + 69 + 95}{103 + 127 + 129 + 139} = 0,588$$

$$w_1 = 0,602, \quad w_2 = 0,528, \quad w_3 = \frac{m_3}{n_3} = \frac{69}{129} = 0,535,$$
$$w_4 = \frac{m_4}{n_4} = \frac{95}{139} = 0,683$$

и подставим значения в формулу (21). Получим
$\chi^2_{\text{набл}} = \dfrac{1}{0,588(1-0,568)} \cdot (103(0,602-0,588)^2 + 127(0,528-0,588)^2 +$
$+ 129(0,535 - 0,588)^2 + 139(0,683 - 0,588)^2) = 8,73.$

По таблице распределения χ^2 по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = r - 1 = 4 - 1 = 3$ находим

$$\chi^2_{\text{кр}} = \chi^2(0,05; 3) = 7,82.$$

Так как $\chi^2_{\text{набл}} = 8,73 > 7,62 = \chi^2_{\text{кр}}$, гипотезу H_0 отвергаем.

Ответ: a) Предполагаемая активность на выборах первой и второй возрастных групп отличаются не значимо; b) отличие долей четырех исследованных групп значимо или существенно при уровне значимости $\alpha = 0,05$.

6.5 Грубые ошибки наблюдений

Рассмотрим частный случай гипотезы о равенстве средних значений двух выборок.

Пусть имеется $n + 1$ измерений некоторой случайной величины x_1, x_2, … x_n, x^*, среди которых значение x^* сильно отличается от остальных и у нас есть основание полагать, что это значение является либо ошибкой эксперимента, неверно записанным показанием наблюдения, либо нетипичным представителем генеральной совокупности. Возникает вопрос, следует ли исключить x^* из выборки, или отличие несущественно. Для ответа на него найдем $\overline{x}_{\text{в}} = \dfrac{1}{n}\sum\limits_{j=1}^{n} x_j$ и сформулируем гипотезу $H_0: \ \overline{x}_{\text{в}} = x^*$ при конкурирующей гипотезе $H_1: \ \overline{x}_{\text{в}} > x^*$ или $\overline{x}_{\text{в}} < x^*$, в зависимости от того, в какую сторону x^* отличается от остальных наблюдаемых значений.

Для проверки гипотезы H_0 построим критерий

$$t = \frac{\overline{x}_{\text{в}} - x^*}{s},$$

где s – "исправленное" среднее квадратическое отклонение. Случайная величина t имеет распределение Стьюдента с $k = n-1$ степенями свободы. Далее находим

$$t_{\text{ набл}} = \frac{\overline{x}_{\text{в}} - x^*}{s} \tag{22}$$

и по таблице распределения Стьюдента по числу степеней свободы $k = n-1$ и уровню значимости α находим $t_{\text{ кр}}$ для односторонней критической области. Если $|t_{\text{ набл}}| < t_{\text{ кр}}$, гипотезу H_0 принимают. Иными словами, нет оснований полагать, что x^* сильно отличается от совокупности остальных наблюдений. При $|t_{\text{ набл}}| > t_{\text{ кр}}$ гипотезу H_0 отвергают.

Пример 11. Эксперимент повторили 9 раз. Получили результаты 253; 264; 272; 312; 308; 293; 253; 267; 337. Найти значение, которое больше всего отличается от среднего по выборке и при уровне значимости $\alpha = 0,025$, проверить, действительно ли это отличие существенно.

Решение. Найдем среднее значение по выборке

$$\frac{253+264+272+312+308+293+253+267+337}{9} = 284,33$$

От этого значения больше всего отличается величина $x^* = 337$. Для проверки гипотезы $H_0: \overline{x}_в = x^*$ найдем $\overline{x}_в = \dfrac{253+264+272+312+308+293+253+267}{8} = 277,75$ и вычислим исправленную дисперсию s^2 по формуле

$$s^2 = \frac{n}{n-1} \cdot D_в = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_в)^2.$$

Тогда $s^2 = \dfrac{1}{7}((253-277,75)^2 + (264-277,75)^2 + (272-277,75)^2 + (312-277,75)^2 + (308-277,75)^2 + +(293-277,75)^2 + (253-277,75)^2 + (267-277,75)^2) = 554,7857$. Отсюда "исправленное" среднее квадратическое отклонение равно: $s = \sqrt{s^2} = \sqrt{554,7857} = 23,55$. Найдем наблюдаемое значение критерия $t_{набл}$ по формуле (22): $t_{\text{ набл}} = \dfrac{\overline{x}_в - x^*}{s} = \dfrac{277,74-337}{23,55} = -2,52$.

По таблице распределения Стьюдента по числу степеней свободы $k = 8-1 = 7$ и уровню значимости $\alpha = 0,025$ находим для односторонней критической области $t_{\text{ кр}} = 2,36$. Так как $|t_{\text{ набл}}| = 2,52 > 2,36 = t_{\text{ кр}}$, гипотезу H_0 отвергаем. То есть значение 337 существенно отличается от остальных значений выборки и его следует отбросить.

Ответ: Значение 337 следует отбросить.

6.6 Гипотеза о значимости выборочного коэффициента корреляции

А. Пусть из нормально распределенной двумерной генеральной совокупности (X, Y) извлечена выборка объема n. По этой выборке найден коэффициент корреляции $r_в$, который является точечной оценкой коэффициента корреляции r генеральной совокупности. Пусть $r_в \neq 0$. Отсюда еще не следует, что коэффициент корреляции в генеральной совокупности r не равен нулю и, следовательно, вопрос о зависимости признаков X и Y остается открытым.

Чтобы ответить на этот вопрос, проверим при заданном уровне значимости α гипотезу $H_0: \ r = 0, \ \ H_1: \ r \neq 0.$

Рассмотрим случайную величину

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}},$$

которая при выполнении гипотезы H_0 имеет распределение Стьюдента с $k = n - 2$ степенями свободы.

Для проверки гипотезы $H_0: \ r = 0, \ \ H_1: \ r \neq 0$ при заданном уровне значимости α, надо вычислить

$$T_{\text{набл}} = \frac{r_{\text{в}}\sqrt{n-2}}{\sqrt{1-r_{\text{в}}^2}} \tag{23}$$

и по таблице распределения Стьюдента, по заданному уровню значимости α и числу степеной свободы $k = n - 2$ найти критическую точку $t_{\text{кр}}(\alpha, k)$ для двусторонней критической области. Если $|T_{\text{набл}}| < t_{\text{кр}}$ — нет оснований отвергать гипотезу H_0. Если $|T_{\text{набл}}| > t_{\text{кр}}$ — гипотезу H_0 отвергают. При этом говорят, что выборочный коэффициент корреляции $r_{\text{в}}$ значимо отличается от нуля.

B. Аналогичным образом проверяется гипотеза о о равенстве нулю генерального коэффициента ранговой корреляции Спирмена $H_0:$ $\rho = 0,$ при конкурарующей гипотезе $H_1: \ \rho \neq 0$ и заданном уровне значимости α. Критерий применим, если объем выборки $n \geq 9$. Для этого надо вычислить

$$T_{\text{набл}} = \frac{\rho_{\text{в}}\sqrt{n-2}}{\sqrt{1-\rho_{\text{в}}^2}} \tag{24}$$

и по таблице распределения Стьюдента, по заданному уровню значимости α и числу степеной свободы $k = n - 2$ найти критическую точку $t_{\text{кр}}(\alpha, k)$ для двусторонней критической области. Если $|T_{\text{набл}}| < t_{\text{кр}}$ — нет оснований отвергать гипотезу H_0. Если $|T_{\text{набл}}| > t_{\text{кр}}$ — гипотезу H_0 отвергают. При этом говорят, что выборочный коэффициентранговой корреляции Спирмена $\rho_{\text{в}}$ значимо отличается от нуля.

Пример 12. Проверить при уравне значимости $\alpha = 0,05$ гипотезу о равенстве нулю коэффициента корреляции, найденного в примере 6 и коэффициента ранговой корреляции примера 7.

Решение. a) В примере 6 $n = 80$ и найдено значение $r_{\text{в}} = -0,80$. Пусть $H_0 : r_{\text{г}} = 0$ и $H_1 : r_{\text{г}} \neq 0$. Тогда, в силу формулы (23) $T_{\text{набл}} = \dfrac{-0,8 \cdot \sqrt{80-2}}{\sqrt{1-0,8^2}} = -11,8$. По таблице распределения Стьюдента для двусторонней критической области находим $t_{\text{кр}} = t(\alpha,\, k) = t(0,05,\, 78) = 1,99$. Так как $|T_{\text{набл}}| = 11,8 > 1,99 = t_{\text{кр}}$, то гипотезу H_0 отвергаем, что означает, что $r_{\text{г}}$ значимо отличается от нуля.

b) В примере 7 $n = 12$ и найдено значение $\rho_{\text{в}} = 0,904$. Здесь $H_0 : \rho_{\text{г}} = 0$ и $H_1 : \rho_{\text{г}} \neq 0$. Тогда из формулы (24) находим $T_{\text{набл}} = \dfrac{0,904 \cdot \sqrt{12-2}}{\sqrt{1-0,904^2}} = 6,7$. По таблице распределения Стьюдента находим $t_{\text{кр}} = t(\alpha,\, k) = t(0,05,\, 10) = 2,23$. Так как $|T_{\text{набл}}| = 6,7 > 2,23 = t_{\text{кр}}$, гипотезу H_0 отвергаем.

Ответ: В примерах 6, 7 коэффициенты $r_{\text{г}}$ и $\rho_{\text{г}}$ значимо отличаются от нуля.

6.7 Гипотезы о законе распределения случайной величины. Критерий χ^2

При исследовании статистических закономерностей бывает важно определить, какому закону подчиняется данная случайная величина или, имея две выборки проверить, они подчиняются одному и тому же закону, или мы имеем по существу разные закономерности.

Одним из методов решения этих задач является применение критерия согласия χ^2 (хи-квадрат) Пирсона. Смысл этого метода в следующем. Пусть сравниваются между собой два ряда. Пусть n_i — **эмпирические частоты** (частоты, полученные в результате наблюдения и пусть n_i' – частоты которые должны быть в предположении, что распределение подчиняется предполагаемому закону (**теоретические частоты**). Например, это может быть нормальное распределение, распределение Пуассона, равномерное распределение или любое другое. Выдвигаеся гипотеза H_0 : { эмпирическое распределение подчиняется предполагаемому закону}. Альтернативная гипотеза H_1 : { закон распределения отличается от теоретического}.

Тогда

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}$$

– случайная величина, с увеличением которой вероятность справедливости гипотезы H_0 уменьшается. Величина χ^2 не зависит от сравниваемого закона распределения и имеет k степеней свободы, где k вычисляется по формуле $k = s-1-r$, где s — число интервалов, r — число параметров предполагаемого закона распределения.

Для проверки гипотезы H_0 при конкурирующей гипотезе H_1 надо вычислить

$$\chi^2_{\text{набл}} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}$$

и по таблице распределения χ^2 по заданному уровню значимости α и числу степеной свободы $k = s - 1 - r$ найти критическую точку $\chi^2_{\text{набл}}(\alpha, k)$. Если $|\chi^2_{\text{набл}}| < \chi^2_{\text{кр}}$ — нет оснований отвергать гипотезу H_0. Если $|\chi^2_{\text{набл}}| > \chi^2_{\text{кр}}$ — гипотезу H_0 отвергают.

Следует отметить, критерий Пирсона применим при объеме выборки $n \geq 50$. Кроме того n_i в каждом интервале должно быть не менее 5. Поэтому при применении критерия малочисленные интервалы объединяют с соседними и только после этого определяют величину s количества полученных интервалов.

Пример 13. Из генеральной совокупности взята выборка. При уровне значимости $\alpha = 0,05$ проверить гипотезу о нормальном распределении признака X в генеральной совокупности.

$[x_{i-1}; x_i)$	[5; 9)	[9; 13)	[13; 17)	[17; 21)	[21; 25)	[25; 29)
n_i	2	3	7	31	28	31

$[x_{i-1}; x_i)$	[29; 33)	[33; 37)	[37; 41)	[41; 45)
n_i	30	12	4	2

Решение.

1. Сначала найдем теоретические частоты из предположения, что распределение признака X подчиняется нормальному закону.

a) Интервалы $[x_{i-1}, x_i)$ заменим на их серединные значения $x_i^* = \dfrac{x_{i-1} + x_i}{2}$ и построим дискретный вариационный ряд:

x_i^*	7	11	15	19	23	27	31	35	39	43	
n_i	2	3	7	31	28	31	30	12	4	2	$n = \sum n_i = 150$

b) Найдем выборочную среднюю и выборочное среднее квадратическое отклонение полученного вариационного ряда по формулам: $\overline{x}_{\text{в}} = \dfrac{\sum x_i^* n_i}{n}$ и $\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = \sqrt{\overline{x^2} - \left(\overline{x}\right)^2}$

Эти вычисления можно упростить, если ввести условные варианты $u_i = \dfrac{x_i - C}{h}$. Тогда

$\overline{u}_{\text{в}} = \dfrac{\sum u_i n_i}{n}$ и $\sigma_{u,\text{в}} = \sqrt{D_{u,\text{в}}} = \sqrt{\overline{u^2} - \left(\overline{u}\right)^2}$. Обратный переход производится по формулам

$$\overline{x}_{\text{в}} = \overline{u}_{\text{в}} \cdot h + C, \qquad \sigma_{\text{в}} = \sigma_{u,\text{в}} \cdot h.$$

В частности при $C = 27$, $h = x_i - x_{i-1} = 4$ получим

x_i^*	7	11	15	19	23	27	31	35	39	43	
u_i	-5	-4	-3	-2	-1	0	1	2	3	4	
n_i	2	3	7	31	28	31	30	12	4	2	
$u_i n_i$	-10	-12	-21	-62	-28	0	30	24	12	8	$\sum u_i x_i = -59$
$u_i^2 n_i$	50	48	63	124	28	0	30	48	36	32	$\sum u_i^2 x_i = 459$

Тогда

$$\overline{u}_{\text{в}} = \frac{-59}{150} = -0,3933, \quad \overline{u^2} = \frac{459}{150} = 3,06,$$

$\sigma_{u,\text{в}} = \sqrt{3,06 - 0,3933^2} = 1,48$. Отсюда

$$\overline{x}_{\text{в}} = -0,3933 \cdot 4 + 27 = 25,43, \qquad \sigma_{\text{в}} = 1,48 \cdot 4 = 5,92.$$

c) Найдем теоретические частоты n'_i нормального распределения $n = 150$ результатов наблюдений, приняв в качестве параметров $a = 25,43$ и $\sigma = 5,92$. Для этого воспользуемся формулами

$$P(\alpha \le X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) \qquad \text{и} \qquad n'_i = n \cdot p_i,$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$ — интегральная функция Лапласа, а $p_i = P(x_{i-1} \le X < x_i)$.

i	$[x_i; x_{i+1})$	$x_i - a$	$x_{i+1} - a$	$z_i = \dfrac{x_i - a}{\sigma}$	$z_{i+1} = \dfrac{x_{i+1} - a}{\sigma}$
1	$5 - 9$	$-$	$-16,43$	$-\infty$	$-2,78$
2	$9 - 13$	$-16,43$	$-12,43$	$-2,78$	$-2,10$
3	$13 - 17$	$-12,43$	$-8,43$	$-2,10$	$-1,42$
4	$17 - 21$	$-8,43$	$-4,43$	$-1,42$	$-0,74$
5	$21 - 25$	$-4,43$	$-0,43$	$-0,74$	$-0,07$
6	$25 - 29$	$-0,43$	$3,57$	$-0,07$	$0,60$
7	$29 - 33$	$3,57$	$7,57$	$0,60$	$1,28$
8	$33 - 37$	$7,57$	$11,57$	$1,28$	$1,95$
9	$37 - 41$	$11,57$	$15,57$	$1,95$	$2,63$
10	$41 - 45$	$15,57$	$-$	$2,63$	∞

Продолжение таблицы

i	$\Phi(z_i)$	$\Phi(z_{i+1})$	$p_i =$ $\Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = np_i$
1	$-0,5$	$-0,4973$	$0,0027$	$0,40$
2	$-0,4973$	$-0,4821$	$0,0152$	$2,28$
3	$-0,4821$	$-0,4222$	$0,0599$	$8,98$
4	$-0,4222$	$-0,2703$	$0,1519$	$22,78$
5	$-0,2703$	$-0,0279$	$0,2424$	$36,36$
6	$-0,0279$	$0,2257$	$0,2536$	$38,04$
7	$0,2257$	$0,3997$	$0,1740$	$26,1$
8	$0,3997$	$0,4744$	$0,0747$	$11,20$
9	$0,4744$	$0,4957$	$0,0213$	$3,20$
10	$0,4957$	$0,5$	$0,0043$	$0,64$
			$\sum = 1$	$\sum = 150$

Замечание. Так как при вычислениях промежуточные результаты округляются, то $\sum p_i$ и $\sum n'_i$ можгут немного отличаться от 1 и n соответственно.

2. Найдем $\chi^2_{\text{набл}} = \sum\limits_{i=1}^{s} \dfrac{(n_i - n'_i)^2}{n'_i}$.

Вычисления оформим в виде таблицы. Но сначала объединим первые три интервала в один, а также три последних. Тогда объединенные интервалы будут содержать не менее 5 элементов: $n_1 = 2 + 3 + 7 = 12$, $n'_1 = 0,4 + 2,28 + 8,99 = 11,67$, $n_6 = 2 + 4 + 12 = 18$, $n'_6 = 11,21 + 3,19 + 0,65 = 15,05$. При этом число интервалов сократится до $s = 6$.

i	n_i		$n'_i = np_i$		$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$
1	2		0,40				
2	3	12	2,28	11,67	0,33	0,11	0,01
3	7		8,99				
4	31		22,78		8,22	67,57	2,97
5	28		36,36		−8,36	69,89	1,92
6	31		38,04		−7,04	49,56	1,30
7	30		26,1		3,9	15,21	0,58
8	12		11,21				
9	4	18	3,19	15,05	2,95	8,70	0,56
10	2		0,65				
							$\chi^2_{\text{набл}} = 7,34$

Число интервалов после объединения стало равным 6: $s = 6$, число оцениваемых параметров $r = 2$. По таблице критических точек распределения χ^2 по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = s-1-r = 6-1-2 = 3$ найдем $\chi^2_{\text{кр}} = \chi^2(0,05;\, 3) = 7,8$. Так как $\chi^2_{\text{набл}} = 7,34 < 7,8 = \chi^2_{\text{кр}}$, то нет оснований отвергать гипотезу H_0.

Ответ: Эмпирические данные согласуются с гипотезой о нормальном распределении генеральной совокупности.

7 Варианты контрольных работ

Задание 1.

По данной выборке 1) найти соответственно наименьшее и наибольшее значения x_{min}, x_{max}; 2) размах вариации; 3) построить вариационный ряд распределения (интервальный и точечный); число интервалов согласовать с формулой Стерджеса; 4) построить полигон или гистограмму а также график накопленных частот; 5) найти эмпирическую функцию $F^*(x)$ и построить ее график.

вар 1. 288; 258; 278; 279; 280; 270; 273; 283; 299; 262; 273; 262; 276; 272; 278; 273; 309; 269; 289; 261; 271; 280; 264; 290; 264; 265;

268; 275; 293; 299; 263; 282; 292; 266; 286; 264; 262; 292; 283; 278; 285; 265; 263; 283; 288; 268; 267; 271; 285; 265.

вар 2. 292; 311; 290; 265; 294; 269; 279; 282; 295; 271; 271; 324; 274; 275; 298; 279; 278; 254; 308; 282; 309; 284; 265; 300; 293; 268; 249; 282; 293; 274; 270; 286; 280; 274; 308; 270; 281; 258; 295; 278; 296; 268; 258; 271; 272; 283; 272; 281; 301; 281

вар 3. 23,1; 37,8; 24,8; 33,6; 29,4; 38,3; 46,6; 51,3; 39,8; 29,2; 30; 38,9; 33,4; 35,4; 23; 42; 26,4; 33,3; 26,1; 30,9; 45,5; 32,7; 34,2; 46,1; 40,6; 37,8; 34,5; 40,9; 38; 35; 37,9; 41,6; 37,2; 36; 33,2; 43; 42,7; 26,8; 27,6; 45,6; 28; 33,1; 41,2; 29,9; 30,4; 34,9; 40,2; 38,5; 38,4; 32,9

вар 4. 23,2; 35,5; 30,9; 34,2; 29,4; 47,7; 43,4; 39,6; 33,5; 24,5; 31,8; 33,2; 27,6; 26,1; 30,7; 27,9; 28,5; 53,3; 28,8; 42,6; 26,4; 30,9; 44,1; 32,8; 29,9; 39,3; 33,1; 38,8; 42,9; 29,5; 33,8; 20,7; 31,7; 44,7; 42,8; 41; 35,3; 31,7; 37,3; 43,5; 18; 31; 25,1; 48,4; 36,3; 43,1; 29,4; 39,5; 38,8; 27,6

вар 5. 26,8; 26,3; 25,5; 30,1; 25,1; 33,8; 28,9; 28,6; 32,7; 28,2; 25,3; 35,2; 29,6; 32,1; 28,2; 23,2; 31,1; 31; 24,4; 25,4; 21,2; 28,4; 27,8; 29,7; 25,5; 28,3; 24,4; 31,6; 33,8; 28,8; 30,1; 25,1; 25,1; 30,3; 22; 26,1; 31,7; 27; 26,6; 27,3; 35,3; 27,4; 29,8; 33,4; 29,1; 31,8; 28,7; 28,8; 25; 24,7

вар 6. 29,4; 31; 29,4; 35,7; 30,8; 31,2; 25; 25,5; 34,8; 29,8; 25,2; 28,4; 34; 28,1; 32,9; 23,4; 35; 28,3; 37,3; 36,7; 33,5; 28,1; 31,8; 35,5; 35,1; 33,7; 33,1; 26,9; 30,7; 32,1; 35,7; 33,9; 29,7; 24,2; 27,6; 23,9; 33,3; 32; 29,8; 33,7; 24,5; 30,8; 33,3; 34,2; 27,8; 32,1; 27,5; 29,6; 28,6; 30,6

вар 7. 28,1; 18,3; 27; 32,9; 34,1; 24,7; 29,8; 27,8; 20,7; 21,7; 29,4; 24,7; 29,7; 28,7; 26,1; 28,8; 23,3; 23,7; 26,5; 20,1; 25,8; 25,4; 27,6; 22,2; 34,1; 26,1; 27; 27,6; 28,3; 25; 28,2; 23,5; 28,6; 18,5; 31,3; 27,4; 23,4; 25,7; 25,4; 18,9; 25,5; 31; 19,7; 28,1; 30,1; 29,7; 24,6; 31,3; 19,9; 19,4

вар 8. 26,5; 29; 27; 26,7; 29,2; 28,7; 33,2; 33,8; 30,3; 30,1; 31,3; 28,7; 25,7; 31,9; 27; 23,4; 37,9; 31,9; 26,5; 23,2; 23,5; 28,1; 30,9; 30,5; 37,4; 24,4; 24,5; 26,2; 28,8; 28,8; 30,2; 28,2; 26,3; 28; 26; 28,2; 24,6; 28,9; 29,4; 29,3; 23,9; 24,9; 28; 25,6; 26,9; 28,2; 30,7; 27,9; 22,3; 31,2

вар 9. 118; 128; 117; 124; 124; 128; 133; 127; 120; 124; 126; 129; 128; 123; 113; 133; 138; 127; 127; 129; 127; 131; 122; 123; 114; 118; 122; 128; 121; 134; 120; 122; 122; 120; 126; 116; 126; 126; 121; 125; 128; 135; 123; 129; 120; 117; 119; 122; 129; 119

вар 10. 116; 119; 116; 108; 119; 116; 123; 118; 109; 125; 118; 117; 127; 115; 121; 126; 126; 119; 122; 125; 118; 122; 122; 119; 126; 118; 121; 119; 119; 116; 115; 112; 112; 115; 125; 114; 119; 126; 114; 110; 118; 123; 115; 119; 118; 115; 123; 120; 129; 117

вар 11. 28; 29,4; 27,1; 28,7; 27,9; 27,7; 27,5; 27,1; 28,2; 29,8; 28,2; 25,9; 28,5; 30,1; 28,1; 27,7; 29,6; 25,3; 30,4; 26,8; 29,3; 27,7; 28,6; 26; 25,9; 29; 29; 27; 29,2; 29,5; 27,6; 28,6; 28,6; 27,4; 29,5; 26,2; 28,5; 28,1; 29,3; 26,4; 27,4; 25,9; 28,8; 29; 27,5; 28,4; 26,7; 27,1; 27,8; 25,9

вар 12. 19,7; 23,9; 22; 22,9; 21,7; 20,7; 23,5; 22,8; 23,4; 24,1; 19; 21,6; 21,8; 24,5; 20,8; 22,5; 22,1; 23,9; 20,8; 20,1; 22,3; 21,1; 22,7; 26,2; 21,7; 21,9; 20; 23,1; 23,4; 21,5; 21,2; 21,1; 22,9; 21,6; 23,3; 21,1; 19,1; 21,2; 22,2; 21,4; 22,1; 23,6; 23,1; 22,2; 22,1; 23,3; 22,9; 21,5; 22,4; 22,2

вар 13. 140; 144; 147; 144; 140; 144; 142; 141; 139; 136; 143; 142; 144; 144; 141; 140; 141; 140; 141; 141; 142; 147; 143; 146; 142; 147; 146; 142; 142; 142; 140; 143; 143; 141; 142; 142; 141; 141; 135; 146; 147; 138; 137; 146; 142; 144; 144; 141; 143; 137

вар 14. 136; 135; 140; 142; 134; 142; 142; 139; 137; 141; 141; 137; 138; 141; 136; 140; 140; 139; 136; 142; 142; 137; 135; 139; 137; 140; 137; 136; 134; 139; 140; 135; 141; 139; 138; 144; 142; 138; 134; 139; 141; 137; 143; 140; 139; 138; 142; 138; 141; 138

вар 15. 52,4; 35; 60,5; 38,7; 53,5; 40,6; 52,3; 52,7; 43,8; 50; 39,1; 26,1; 37,9; 58,3; 44,3; 41,4; 54,1; 48,5; 54,8; 45,5; 31,6; 62,9; 39,4; 38,9; 49,5; 49,8; 43,1; 46,8; 52,3; 41,2; 34,1; 42,8; 33,7; 62; 43,9; 55,7; 35,9; 26,7; 47,6; 40,3; 46,3; 29,4; 49,8; 52; 57,3; 47,8; 42,4; 45,1; 50,4; 29,9

вар 16. 44,7; 45,9; 54,7; 50,8; 48,1; 35,9; 59,3; 36,9; 45,1; 46,4; 56,6; 54,7; 42,2; 48; 42; 40,7; 44,8; 45,4; 30; 51,8; 33,7; 46,7; 32; 43; 48,7; 52,1; 39; 34,4; 38,2; 56; 38,5; 52; 59,2; 52,1; 31,4; 34; 46,4; 50,5; 50,4; 30,4; 49,2; 38,7; 49,4; 48,8; 50,9; 40,5; 45,3; 44,5; 40,3; 55,8

вар 17. 39,2; 37,2; 31; 29,7; 46,6; 29,9; 29,7; 38,2; 32,3; 43,1; 27,9; 43; 45,3; 32,4; 39,9; 35,2; 35; 25,5; 36,3; 31,1; 38,7; 27,6; 24,4; 30,6; 37,1; 37,3; 29,2; 33,3; 36,9; 34,7; 34,7; 41,4; 40,3; 35; 28,3; 37,2; 31,8; 41,1; 39,3; 46,1; 30,9; 38; 29,3; 20,5; 33,8; 30,3; 26,4; 31,9; 31,9; 30,1

вар 18. 29,2; 40,8; 31,4; 39,3; 42; 38,2; 40,1; 41,2; 48,8; 44,2; 38,7; 52,1; 31,7; 47,8; 47,8; 37,5; 38; 40; 50,3; 52; 32,2; 41,2; 38; 36,7; 37,1;

44,8; 38,2; 28,1; 33,2; 41,9; 39; 39,9; 28,2; 36,4; 43,6; 35; 29,2; 36,9; 41,8; 53,2; 32,8; 36,7; 26,3; 38,8; 34,5; 46,2; 34,2; 35,5; 35,5; 31

вар 19. 64,4; 54,4; 54,8; 50,1; 47,8; 52; 60,4; 57,9; 61,8; 56,4; 58,7; 47,2; 45,5; 46,4; 55,2; 44,5; 56,5; 53,8; 53,7; 52,2; 50,9; 59,9; 55,3; 45,9; 47,9; 62,7; 48,2; 63,2; 59,5; 57,7; 44,8; 44,2; 56,5; 46; 50,9; 45,9; 63,1; 51,5; 52,7; 51,1; 62,4; 55,3; 52,4; 61,1; 56,6; 55,2; 56,9; 49,6; 60,4; 60,3

вар 20. 52,4; 51,1; 38,3; 49,7; 50,6; 46,2; 54,9; 48,2; 50,8; 44,3; 57,2; 46; 60,5; 60,7; 51,7; 49,3; 60,8; 45,6; 55,1; 46,3; 40,4; 55; 52; 57,9; 51,1; 49,5; 55; 44,9; 60; 45,6; 51,5; 58,3; 56,2; 58,5; 46,2; 58,6; 55; 55,2; 50,8; 53,9; 55,6; 44,6; 46,2; 44,2; 55,7; 42,9; 48,2; 53,1; 55,1; 44,9

Задание 2.

Для вариационного ряда, построенного в задании 1, найти: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение; 2) несмещенную оценку генеральной дисперсии, а также «исправленное» среднее квадратическое отклонение; 3) медиану и моду(интервального ряда); 4) коэффициент вариации; 5) асимметрию и эксцесс.

Задание 3.

Из генеральной совокупности, подчиненной нормальному закону распределения взяты три независимые выборки объема n_i. По этим выборкам найдены выборочные средние и выборочные дисперсии. Выборки объединены в одну. Найти $\overline{x}_{\text{в}}$ и $D_{\text{в}}$ объединенной выборки.

	n_1	$\overline{x}_{\text{в}1}$	$D_{\text{в}1}$	n_2	$\overline{x}_{\text{в}2}$	$D_{\text{в}2}$	n_3	$\overline{x}_{\text{в}3}$	$D_{\text{в}3}$
вар 1.	9	147	12	13	124	11	7	135	6
вар 2.	15	34,3	4,0	7	36,1	3,2	12	37,9	2,2
вар 3.	21	9,26	1,42	11	7,11	2,07	10	10,20	1,85
вар 4.	23	31,2	2,9	25	27,4	3,2	19	25,2	2,4
вар 5.	7	790	75	22	778	47	18	781	54
вар 6.	23	54,9	1,9	31	51,8	0,7	16	57,2	2,1
вар 7.	14	131	15	21	129	18	25	125	13

вар 8.	42	47,0	3,1	16	45,6	2,8	18	46,2	2,5
вар 9.	17	154	45	37	147	29	27	144	42
вар 10.	22	32,6	2,4	26	34,1	2,7	13	31,2	3,5
вар 11.	24	31,2	2,5	35	34,4	2,9	10	29,5	3,3
вар 12.	17	179	15	32	178	14	18	181	21
вар 13.	32	564	29	10	583	6	26	572	24
вар 14.	14	19,7	6,2	21	18,1	3,8	16	17,5	1,9
вар 15.	22	47,3	3,4	18	45,9	2,9	13	46,8	2,4
вар 16.	27	154	41	31	149	37	17	147	42
вар 17.	12	23,6	2,9	24	24,1	2,7	19	23,2	3,5
вар 18.	21	35,2	3,2	35	34,9	2,9	18	37,5	3,3
вар 19.	24	17,9	1,9	5	17,1	1,4	19	18,4	2,3
вар 20.	27	412	13	18	407	1,4	7	418	2,3

Задание 4.

Из генеральной совокупности большого объема, подчиненной нормальному закону распределения взята выборка объема n. По этой выборке найдены выборочная средняя и выборочная дисперсия. Найти доверительные интервалы для оценки с надежностью $\gamma = 0,99$ генеральной средней $\overline{x}_{\text{г}}$ и генерального среднего квадратического отклонения σ.

	n	$\overline{x}_{\text{в}}$	$D_{\text{в}}$		n	$\overline{x}_{\text{в}}$	$D_{\text{в}}$
вар 1.	9	147	12	**вар 2.**	15	34,3	4,0
вар 3.	21	9,26	1,42	**вар 4.**	23	31,2	2,9
вар 5.	7	790	75	**вар 6.**	23	54,9	1,9
вар 7.	14	131	15	**вар 8.**	24	47,0	3,1
вар 9.	17	154	45	**вар 10.**	22	32,6	2,4
вар 11.	24	31,2	2,5	**вар 12.**	17	179	15
вар 13.	22	564	29	**вар 14.**	14	19,7	6,2
вар 15.	22	47,3	3,4	**вар 16.**	27	154	41
вар 17.	12	23,6	2,9	**вар 18.**	21	35,2	3,2

вар 19.	24	17,9	1,9	**вар 20.**	19	126	13

Задание 5.

Найти минимальный объем выборки, при котором с надежностью γ точность оценки математического ожидания a генеральной совокупности по выборочной средней равна δ, если известно среднее квадратическое отклонение σ нормально распределенной генеральной совокупности.

	γ	δ	σ		γ	δ	σ
вар 1.	0,975	0,2	1,2	**вар 2.**	0,925	0,3	1,4
вар 3.	0,945	5,4	17,4	**вар 4.**	0,823	31,2	110
вар 5.	0,994	5,7	12,2	**вар 6.**	0,932	36,5	141
вар 7.	0,916	3,7	16,7	**вар 8.**	0,845	13,2	71
вар 9.	0,935	7,3	23,7	**вар 10.**	0,851	23,9	132
вар 11.	0,935	3,2	14,8	**вар 12.**	0,828	37,4	146
вар 13.	0,943	5,3	19,8	**вар 14.**	0,878	33,7	129
вар 15.	0,922	3,8	21,7	**вар 16.**	0,810	43,1	137
вар 17.	0,945	7,6	27,3	**вар 18.**	0,921	26,3	154
вар 19.	0,915	9,2	28,4	**вар 20.**	0,875	24,3	115

Задание 6.

По выборке, извлеченной из двумерной нормальной генеральной совокупности (X, Y), составлена корреляционная таблица. Найти выборочный коэффициент корреляции. Написать уравнение регрессии Y на X. Проверить гипотезу о значимости коэффициента корреляции.

вар1.

Y \ X	2,5	2,8	3,1	3,4	n_y
11	11	21	–	–	32
14	–	13	26	–	39
17	–	–	21	13	34
n_x	11	34	47	13	$n = 105$

вар2.

Y \ X	26	29	32	35	n_y
8	9	19	–	–	28
12	–	13	27	–	40
16	–	–	19	13	32
n_x	9	32	46	13	$n = 100$

вар3.

Y \ X	124	127	130	133	n_y
2,3	–	–	7	21	28
2,9	–	2	26	10	38
3,5	15	19	–	–	34
n_x	15	21	33	31	$n = 100$

вар4.

Y \ X	11	17	23	29	n_y
51	–	–	6	18	24
55	–	8	21	4	33
59	12	16	–	–	28
n_x	12	24	27	22	$n = 85$

вар5. X

Y	3,2	3,8	4,4	5,0	n_y
11	9	23	–	–	32
15	–	15	22	–	37
19	–	–	19	12	31
n_x	9	38	41	12	$n = 100$

вар6. X

Y	17	28	39	50	n_y
6	5	7	–	–	12
13	–	8	9	4	21
20	–	2	12	13	27
n_x	5	17	21	17	$n = 60$

вар7. X

Y	25	26,5	28	29,5	n_y
15	3	8	–	–	11
19	–	18	12	–	30
23	–	–	11	13	24
n_x	3	26	23	13	$n = 65$

вар8. X

Y	3,2	3,8	4,4	5,0	n_y
7	3	6	–	1	10
15	2	9	12	–	23
23	–	1	8	13	22
n_x	5	16	20	14	$n = 55$

вар9.

Y \ X	$2,5$	$6,5$	$10,5$	$14,5$	n_y
$1,5$	–	1	11	17	29
$1,8$	–	17	20	1	38
$2,1$	6	19	2	1	28
n_x	6	37	33	19	$n=95$

вар10.

Y \ X	140	170	200	230	n_y
$0,7$	1	1	6	5	13
$1,3$	–	8	12	2	22
$1,9$	7	7	1	–	15
n_x	8	16	19	7	$n=50$

вар11.

Y \ X	12	16	20	24	n_y
$5,7$	8	7	1	–	16
$6,0$	–	8	9	1	18
$6,3$	–	–	10	6	16
n_x	8	15	20	7	$n=50$

вар12.

Y \ X	14	18	22	26	n_y
$13,4$	–	–	13	23	36
$13,7$	–	18	14	–	32
$14,0$	21	11	–	–	32
n_x	21	29	27	23	$n=100$

вар13. X

Y	$1,25$	$1,28$	$1,31$	$1,34$	n_y
5	11	9	–	1	21
11	1	12	13	–	26
17	–	–	11	12	23
n_x	12	21	24	13	$n = 70$

вар14. X

Y	25	28	31	34	n_y
17	11	21	–	–	32
22	–	13	26	–	39
27	–	–	21	13	34
n_x	11	34	47	13	$n = 105$

вар15. X

Y	100	106	112	118	n_y
24	8	10	–	–	18
31	–	12	11	1	24
38	1	–	14	13	28
n_x	9	22	25	14	$n = 70$

вар16. X

Y	44	50	56	62	n_y
$9,1$	–	–	13	11	24
$9,8$	–	15	10	–	25
$10,5$	17	9	–	–	26
n_x	17	24	32	11	$n = 75$

вар17.

Y \ X	1,2	1,7	2,2	2,7	n_y
17	4	6	1	–	11
21	–	–	8	5	13
25	–	–	10	6	16
n_x	4	6	19	11	$n = 40$

вар18.

Y \ X	42	51	60	69	n_y
0,03	–	–	15	21	36
0,09	–	16	12	–	28
0,15	23	13	–	–	36
n_x	23	29	27	21	$n = 100$

вар19.

Y \ X	2,5	2,9	3,3	3,7	n_y
27	–	–	14	13	27
37	–	15	12	–	27
47	10	11	–	–	21
n_x	10	26	26	13	$n = 75$

вар20.

Y \ X	0,2	0,5	0,8	1,1	n_y
62	1	4	42	21	68
68	–	30	23	–	53
74	45	34	–	–	79
n_x	46	68	65	21	$n = 200$

Задание 7.

По двум независимым выборкам, объемы которых n_1 и n_2 извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии s_X^2 и s_Y^2. При уровне значимости α проверить гипотезу $H_0 : D(X) = D(Y)$ при конкурирующей гипотезе H_1.

	n_1	n_2	s_X^2	s_Y^2	α	H_1
вар 1.	10	12	0,75	0,34	0,05	$D_X > D_Y$
вар 2.	11	12	1,24	0,82	0,05	$D_X > D_Y$
вар 3.	12	10	7,4	12,2	0,1	$D_X \neq D_Y$
вар 4.	10	9	2,8	6,22	0,01	$D_X < D_Y$
вар 5.	11	12	4,6	7,35	0,01	$D_X < D_Y$
вар 6.	8	12	1,74	5,31	0,02	$D_X \neq D_Y$
вар 7.	17	12	3,21	1,35	0,05	$D_X > D_Y$
вар 8.	11	12	14,4	6,2	0,05	$D_X > D_Y$
вар 9.	15	12	7,2	13,6	0,1	$D_X \neq D_Y$
вар 10.	10	12	23	65	0,01	$D_X < D_Y$
вар 11.	9	12	34	67	0,01	$D_X < D_Y$
вар 12.	12	8	1,62	4,71	0,02	$D_X \neq D_Y$
вар 13.	10	12	25,2	71,1	0,05	$D_X < D_Y$
вар 14.	11	12	11,4	5,2	0,05	$D_X > D_Y$
вар 15.	12	10	17,4	39,2	0,1	$D_X \neq D_Y$
вар 16.	10	9	12,8	35,3	0,01	$D_X < D_Y$
вар 17.	11	12	34,3	7,23	0,01	$D_X > D_Y$
вар 18.	8	12	210	71,5	0,02	$D_X \neq D_Y$
вар 19.	11	9	32,8	95,3	0,01	$D_X < D_Y$
вар 20.	11	12	24,1	63,2	0,01	$D_X > D_Y$

Задание 8.

Задача 1. Партия товара принимается к реализации, если количество бракованных изделий в ней не превосходит $P\%$. По выборке объема n брак составил m изделий. Проверить при уровне значимости α, следует ли принять эту партию к реализации?

Вариант 1- 10. Решить задачу 1 при следущих численных значениях данных

	$P\%$	n	m	α
вар 1.	3	400	18	0,05
вар 2.	5	350	20	0,02
вар 3.	4	250	16	0,01
вар 4.	3	350	13	0,03
вар 5.	4	300	18	0,04
вар 6.	10	180	25	0,02
вар 7.	2,5	400	18	0,01
вар 8.	4	300	19	0,03
вар 9.	3	600	25	0,05
вар 10.	3	250	13	0,01

Задача 2. Производитель утверждает,что количество стандартных изделий составляет не менее $P\%$. Для контороля качества было проверено n изделий, среди которых оказалось m стандартных. Следует ли согласиться с утверждением производителя при уровне значимости α?

Вариант 11-20. Решить задачу 2 при следущих численных значениях данных

	$P\%$	n	m	α
вар 11.	95	400	370	0,015
вар 12.	97	350	335	0,05
вар 13.	90	250	215	0,02
вар 14.	85	350	310	0,01
вар 15.	96	300	282	0,03
вар 16.	90	180	155	0,04
вар 17.	96	475	447	0,02
вар 18.	95	300	278	0,01
вар 19.	92	540	482	0,03
вар 20.	97	450	432	0,05

Задание 9.

Даны результаты 9 наблюдений. Найти значение, которое больше всего отличается от среднего по выборке и при уровне значимости α проверить, действительно ли это отличие существенно.

вар 1. 54,4; 53,5; 54,1; 52,3; 54,8; 53,5; 53,1; 52,3; 51,3. $\alpha = 0,025.$

вар 2. 25,9; 26,2; 24,3; 21,9; 22,3; 26,1; 22,4; 26,3; 28,6. $\alpha = 0,05.$

вар 3. 116,5; 119,7; 115,2; 118,2; 117,5; 116,9; 122,9; 121,9; 112,3.
$\alpha = 0,025.$

вар 4. 34,1; 32,8; 33,1; 31,3; 36,9; 31,9; 30,9; 36,2; 38,2. $\alpha = 0,05.$

вар 5. 25,3; 25; 25,1; 23,4; 25,9; 25,2; 27; 27,2; 29,3. $\alpha = 0,025.$

вар 6. 27,1; 29,7; 23,6; 26,2; 26,1; 30,2; 20,1; 24,1; 17,4. $\alpha = 0,025.$

вар 7. 35,5; 35,8; 37,5; 34,2; 32,9; 32,1; 34,7; 35,3; 39,3. $\alpha = 0,05.$

вар 8. 73,7; 72,2; 71,9; 71,7; 72,2; 75,1; 75; 72,6; 76,1. $\alpha = 0,025.$

вар 9. 129,2; 116,9; 131,9; 123,4; 128,7; 116,6; 127,7; 123,5; 110,4.
$\alpha = 0,025.$

вар 10. 130,5; 129,4; 129; 132,5; 127; 127,2; 128,9; 132,7; 135,8.
$\alpha = 0,05.$

вар 11. 18; 18,1; 16,4; 17,3; 19; 16,8; 16,4; 16,9; 14,1. $\alpha = 0,05.$

вар 12. 12,8; 12,8; 13,4; 13; 13,3; 13,1; 13,3; 13; 13,9. $\alpha = 0,025.$

вар 13. 37; 37,4; 37,6; 33,5; 37; 35,3; 37,8; 35,4; 32,1. $\alpha = 0,05.$

вар 14. 55,4; 53,4; 56,2; 52,5; 54,1; 53; 53,4; 54,1; 51,1. $\alpha = 0,025.$

вар 15. 70,2; 71,6; 66,8; 74; 73,7; 75,9; 68,6; 72,3; 78,1. $\alpha = 0,025$.

вар 16. 39,2; 33,7; 38,4; 38,9; 36,7; 38,5; 39,9; 36,1; 32,3. $\alpha = 0,05$.

вар 17. 36,9; 38,9; 40; 44,8; 41,3; 42; 41,4; 40; 48,4. $\alpha = 0,025$.

вар 18. 56,7; 54,9; 58; 57,1; 54,4; 53,8; 55,2; 56,7; 52,2. $\alpha = 0,05$.

вар 19. 37,7; 37,2; 35,2; 38,6; 43,2; 45,5; 43,2; 43,8; 48,2. $\alpha = 0,025$.

вар 20. 40,3; 42,1; 45,3; 40,8; 45,3; 42,7; 45,1; 47; 48,9. $\alpha = 0,025$.

Задание 10.

Два преподавателя оценили знания двенадцати учащихся по стобальной системе и выставили им следующие оценки (в первой строке указано количество баллов, выставленных первым преподавателем, а во второй — вторым): А) Найти выборочный коэффициент ранговой корреляции Спирмена между оценками двух преподавателей. В) При уровне значимости $\alpha = 0,05$ проверить гипотезу о равенстве нулю генерального коэффициента ранговой корреляции Спирмена. Другими словами, требуется проверить, является ли значимой ранговая корреляционная связь между оценками двух преподавателей.

вар 1.	98	94	88	61	94	70	63	61	60	61	56	51
	99	91	91	74	78	91	64	91	52	53	48	62
вар 2.	96	92	78	96	76	70	62	70	60	58	56	70
	99	91	93	93	78	65	66	66	52	66	48	62
вар 3.	98	95	78	96	76	95	62	76	60	58	56	76
	99	91	91	93	78	65	65	66	52	52	48	63
вар 4.	93	95	77	95	76	77	62	70	77	58	56	70
	99	94	93	94	78	63	66	66	54	66	52	62

вар 5.	43	45	67	55	76	77	62	70	77	58	56	70
	64	64	93	94	78	78	66	62	54	66	52	62
вар 6.	92	95	87	95	76	76	64	70	77	68	56	70
	89	94	93	54	78	63	68	66	54	68	52	62
вар 7.	83	85	71	95	71	76	62	71	70	58	52	70
	94	94	93	94	77	63	66	66	54	66	52	62
вар 8.	96	92	78	96	76	70	62	70	60	58	56	70
	98	94	88	61	94	70	63	61	60	61	56	51
вар 9.	99	91	91	74	78	91	64	91	52	53	48	62
	99	91	93	93	78	65	66	66	52	66	48	62
вар 10.	99	94	93	94	78	63	66	66	54	66	52	62
	98	95	78	96	76	95	62	76	60	58	56	76
вар 11.	99	91	91	93	78	65	65	66	52	52	48	63
	99	94	93	94	78	63	66	66	54	66	52	62
вар 12.	89	94	93	54	78	63	68	66	54	68	52	62
	64	64	93	94	78	78	66	62	54	66	52	62
вар 13.	92	95	87	95	76	76	64	70	77	68	56	70
	43	45	67	55	76	77	62	70	77	58	56	70
вар 14.	94	94	93	94	77	63	66	66	54	66	52	62
	43	35	41	41	51	76	51	71	52	58	52	70
вар 15.	77	62	77	75	79	52	64	71	52	45	48	61
	84	56	93	93	59	64	66	64	52	64	58	47
вар 16.	94	93	93	94	81	65	63	65	65	66	57	59
	97	94	83	91	75	91	68	75	63	75	59	62
вар 17.	93	95	77	95	76	77	62	70	77	58	56	70
	95	97	96	94	82	66	82	66	57	66	51	02
вар 18.	81	84	92	65	75	65	63	67	53	65	58	61
	71	66	92	95	77	71	65	71	54	65	50	62
вар 19.	91	96	85	96	73	76	73	77	73	69	61	61
	49	47	77	59	61	77	60	71	77	53	61	72

вар 20.	91	97	91	95	79	65	79	61	54	65	42	61
	49	37	47	49	47	74	52	74	52	59	52	73

Задание 11.

Используя критерий Пирсона, при уровне значимости α проверить, А) согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки (x_i – первая строка, соответствующие частоты n_i – вторая строка); В) согласуется ли гипотеза о равномерном распределении генеральной совокупности X с эмпирическим распределением выборки (x_i – первая строка, соответствующие частоты m_i – третья строка)

вар 1. $\alpha = 0,025$.

$[x_{i-1}; x_i)$	[8; 11)	[11; 14)	[14; 17)	[17; 20)	[20; 23)
n_i	2	5	12	21	37
m_i	13	17	23	21	14

$[x_{i-1}; x_i)$	[23; 26)	[26; 29)	[29; 32)	
n_i	21	14	8	$\sum n_i = 120$
m_i	25	23	4	$\sum m_i = 140$

вар 2. $\alpha = 0,01$.

$[x_{i-1}; x_i)$	[13; 17)	[17; 21)	[21; 25)	[25; 29)	[29; 33)	[33; 37)
n_i	2	2	5	19	28	20
m_i	4	18	13	26	12	17

$[x_{i-1}; x_i)$	[37; 41)	[41; 45)	
n_i	18	6	$\sum n_i = 100$
m_i	5	19	$\sum m_i = 114$

вар 3. $\alpha = 0,005$.

$[x_{i-1}; x_i)$	[21; 28)	[28; 35)	[35; 42)	[42; 49)	[49; 56)	[56; 63)
n_i	4	5	16	19	22	17
m_i	6	19	15	31	11	8

$[x_{i-1}; x_i)$	[63; 70)	[70; 77)	
n_i	9	8	$\sum n_i = 100$
m_i	17	19	$\sum m_i = 126$

вар 4. $\alpha = 0,025.$

$[x_{i-1};\, x_i)$	$[23;\, 31)$	$[31;\, 39)$	$[39;\, 47)$	$[47;\, 55)$	$[55;\, 63)$	$[63;\, 71)$
n_i	2	8	11	17	21	34
m_i	3	18	15	21	15	18

$[x_{i-1};\, x_i)$	$[71;\, 79)$	$[79;\, 87)$	
n_i	19	5	$\sum n_i = 117$
m_i	4	20	$\sum m_i = 114$

вар 5. $\alpha = 0,01.$

$[x_{i-1};\, x_i)$	$[3;\, 3,5)$	$[3,5;\, 4)$	$[4;\, 4,5)$	$[4,5;\, 5)$	$[5;\, 5,5)$	$[5,5;\, 6)$
n_i	6	8	19	25	21	17
m_i	19	21	19	17	19	27

$[x_{i-1};\, x_i)$	$[6;\, 6,5)$	$[6,5;\, 7)$	
n_i	3	1	$\sum n_i = 100$
m_i	5	6	$\sum m_i = 133$

вар 6. $\alpha = 0,005.$

$[x_{i-1};\, x_i)$	$[13;\, 17)$	$[17;\, 21)$	$[21;\, 25)$	$[25;\, 29)$	$[29;\, 33)$	$[33;\, 37)$
n_i	2	2	5	17	21	14
m_i	3	9	8	19	12	13

$[x_{i-1};\, x_i)$	$[37;\, 41)$	$[41;\, 45)$	
n_i	5	4	$\sum n_i = 70$
m_i	4	9	$\sum m_i = 77$

вар 7. $\alpha = 0,025.$

$[x_{i-1};\, x_i)$	$[1;\, 2,5)$	$[2,5;\, 4)$	$[4;\, 5,5)$	$[5,5;\, 7)$	$[7;\, 8,5)$	$[8,5;\, 10)$
n_i	1	2	5	23	19	14
m_i	27	21	16	29	17	18

$[x_{i-1};\, x_i)$	$[10;\, 11,5)$	$[11,5;\, 13)$	
n_i	8	3	$\sum n_i = 95$
m_i	27	21	$\sum m_i = 176$

вар 8. $\alpha = 0,001.$

$[x_{i-1};\, x_i)$	$[24;\, 32)$	$[32;\, 40)$	$[40;\, 48)$	$[48;\, 56)$	$[56;\, 64)$	$[64;\, 72)$
n_i	3	9	17	24	16	13
m_i	6	19	9	14	18	6

$[x_{i-1};\, x_i)$	$[72;\, 80)$	$[80;\, 88)$	
n_i	6	2	$\sum n_i = 90$
m_i	17	13	$\sum m_i = 102$

вар 9. $\alpha = 0,01.$

$[x_{i-1};\, x_i)$	$[11;\, 18)$	$[18;\, 25)$	$[25;\, 32)$	$[32;\, 39)$	$[39;\, 46)$	$[46;\, 53)$
n_i	2	4	12	19	23	10
m_i	9	19	13	14	21	12

$[x_{i-1};\, x_i)$	$[53;\, 60)$	$[60;\, 67)$	
n_i	3	2	$\sum n_i = 75$
m_i	3	7	$\sum m_i = 98$

вар 10. $\alpha = 0,005.$

$[x_{i-1};\, x_i)$	$[3;\, 14)$	$[14;\, 25)$	$[25;\, 36)$	$[36;\, 47)$	$[47;\, 58)$	$[58;\, 69)$
n_i	1	4	8	13	19	8
m_i	15	7	13	24	16	9

$[x_{i-1};\, x_i)$	$[69;\, 80)$	$[80;\, 91)$	
n_i	5	2	$\sum n_i = 60$
m_i	4	2	$\sum m_i = 85$

вар 11. $\alpha = 0,025.$

$[x_{i-1};\, x_i)$	$[23;\, 28)$	$[28;\, 33)$	$[33;\, 38)$	$[38;\, 43)$	$[43;\, 48)$	$[48;\, 53)$
n_i	3	6	10	18	10	7
m_i	13	5	19	15	10	14

$[x_{i-1};\, x_i)$	$[53;\, 58)$	$[58;\, 63)$	
n_i	5	1	$\sum n_i = 60$
m_i	10	4	$\sum m_i = 90$

вар 12. $\alpha = 0,001.$

$[x_{i-1};\, x_i)$	$[7;\, 13)$	$[13;\, 19)$	$[19;\, 25)$	$[25;\, 31)$	$[31;\, 37)$	$[37;\, 43)$
n_i	4	6	21	25	19	9
m_i	8	4	10	12	10	4

$[x_{i-1};\, x_i)$	$[43;\, 49)$	$[49;\, 55)$	
n_i	5	1	$\sum n_i = 90$
m_i	12	5	$\sum m_i = 65$

вар 13. $\alpha = 0,01.$

$[x_{i-1};\, x_i)$	$[0,5;\, 2)$	$[2;\, 3,5)$	$[3,5;\, 5)$	$[5;\, 6,5)$	$[6,5;\, 8)$	$[8;\, 9,5)$
n_i	3	6	8	21	19	16
m_i	14	17	4	19	17	6

$[x_{i-1};\, x_i)$	$[9,5;\, 11)$	$[11;\, 12,5)$	
n_i	5	2	$\sum n_i = 80$
m_i	7	6	$\sum m_i = 90$

вар 14. $\alpha = 0,005.$

$[x_{i-1};\, x_i)$	$[8;\, 15)$	$[15;\, 22)$	$[22;\, 29)$	$[29;\, 36)$	$[36;\, 43)$	$[43;\, 50)$
n_i	3	7	15	19	24	7
m_i	14	4	10	13	9	13

$[x_{i-1};\, x_i)$	$[50;\, 57)$	$[57;\, 64)$	
n_i	4	1	$\sum n_i = 80$
m_i	3	6	$\sum m_i = 72$

вар 15. $\alpha = 0,025.$

$[x_{i-1};\, x_i)$	$[15;\, 18)$	$[18;\, 21)$	$[21;\, 24)$	$[24;\, 27)$	$[27;\, 30)$	$[30;\, 33)$
n_i	4	6	12	16	14	8
m_i	11	9	13	11	17	4

$[x_{i-1};\, x_i)$	$[33;\, 36)$	$[36;\, 39)$	
n_i	3	2	$\sum n_i = 65$
m_i	11	8	$\sum m_i = 84$

вар 16. $\alpha = 0,001.$

$[x_{i-1};\, x_i)$	$[13;\, 22)$	$[22;\, 31)$	$[31;\, 40)$	$[40;\, 49)$	$[49;\, 58)$	$[58;\, 67)$
n_i	2	3	7	20	14	8
m_i	7	3	12	19	13	8

$[x_{i-1};\, x_i)$	$[67;\, 76)$	$[76;\, 85)$	
n_i	5	1	$\sum n_i = 60$
m_i	17	11	$\sum m_i = 90$

вар 17. $\alpha = 0,01.$

$[x_{i-1};\, x_i)$	$[2;\, 2,5)$	$[2,5;\, 3)$	$[3;\, 3,5)$	$[3,5;\, 4)$	$[4;\, 4,5)$	$[4,5;\, 5)$
n_i	1	3	10	13	15	10
m_i	14	4	14	15	12	5

$[x_{i-1};\, x_i)$	$[5;\, 5,5)$	$[5,5;\, 6)$	
n_i	5	2	$\sum n_i = 60$
m_i	20	11	$\sum m_i = 90$

вар 18. $\alpha = 0,005.$

$[x_{i-1};\, x_i)$	$[3;\, 15)$	$[15;\, 27)$	$[27;\, 39)$	$[39;\, 51)$	$[51;\, 63)$	$[63;\, 75)$
n_i	4	13	19	27	22	12
m_i	15	14	10	15	11	6

$[x_{i-1};\, x_i)$	$[75;\, 87)$	$[87;\, 99)$	
n_i	6	2	$\sum n_i = 105$
m_i	10	17	$\sum m_i = 98$

вар 19. $\alpha = 0,025.$

$[x_{i-1};\, x_i)$	$[11;\, 18)$	$[18;\, 25)$	$[25;\, 32)$	$[32;\, 39)$	$[39;\, 46)$	$[46;\, 53)$
n_i	2	5	13	19	22	13
m_i	14	10	5	18	15	13

$[x_{i-1};\, x_i)$	$[53;\, 60)$	$[60;\, 67)$	
n_i	4	2	$\sum n_i = 80$
m_i	19	11	$\sum m_i = 105$

вар 20. $\alpha = 0,001.$

$[x_{i-1};\, x_i)$	$[8;\, 11)$	$[11;\, 14)$	$[14;\, 17)$	$[17;\, 20)$	$[20;\, 23)$	$[23;\, 26)$
n_i	1	4	8	18	23	12
m_i	9	6	10	13	8	13

$[x_{i-1};\, x_i)$	$[26;\, 29)$	$[29;\, 32)$	
n_i	6	3	$\sum n_i = 75$
m_i	13	12	$\sum m_i = 84$

Список литературы

1. Вентцель Е.С. Теория вероятностей. М., Высшая школа, 2002.

2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М., Высшая школа, 2004.

3. Гмурман В.Е. Теория вероятностей и математическая статистика. М., Высшая школа, 2003.

4. Емельянов Г.В., Скитович В.П. Задачник по теории вероятностей и математической статистике. М., Изд. МГУ, 1967.

5. Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика. М, Высшая школа, 1982.

6. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. М., Наука, 1978.

7. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. Под ред. А.А.Свешникова. М.,Наука, 1970.

8. Кремер Н. Ш. Теория вероятностей и математическая статистика: учебник и практикум для академического бакалавриата / Н.Ш. Кремер.- 4-е изд., пер. и доп. – М.: Юрайт, 2015.

9. Ковалев Е.А. Теория вероятностей и математическая статистика для экономистов: учебник и практикум для бакалавриата и магистратуры / Е.А. Ковалев, Г.А. Медведев; под общ. ред. Г.А. Медведева . – 2-е изд., испр. и доп. –М.: Издательство Юрайт, 2017. –284 с. – Серия: Бакалавр и магистр. Академический курс.

Содержание

Учебное издание

Авторы-составители

Ирисов Андрей Егорович
Сметанин Юрий Михайлович

ОПИСАТЕЛЬНАЯ СТАТИСТИКА И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Учебно-методическое пособие

Авторская редакция

Отпечатано с оригинал-макета заказчика

Подписано в печать 16.04.2021. Формат 60x84 $^1/_{16}$.
Усл. печ. л. 3,5. Уч.-изд. л. 3,1.
Тираж 300 экз. Заказ № 779.

Издательский центр «Удмуртский университет»
426034, Ижевск, ул. Университетская, 1, корп. 4, каб. 207.
Тел./факс: (3412) 500-295 E-mail: editorial@udsu.ru

Типография Издательского центра
«Удмуртский университет»
426034, Ижевск, ул. Университетская, 1, корп. 2.
Тел. 68-57-18