|  |  |  |
| --- | --- | --- |
|  |  |  |

**Далее:** [§39. Математическая статистика в](http://cito-web.yspu.org/link1/metod/theory/node45.html) **Вверх:** [Глава V. Математическая статистика](http://cito-web.yspu.org/link1/metod/theory/node37.html) **Назад:** [§37. Статистические гипотезы](http://cito-web.yspu.org/link1/metod/theory/node43.html) 

**§38. Критерии Пирсона и Стьюдента**

Рассмотрим нормальные независимые случайные величины $X_{i}(i = 1, 2, \ldots
, n)$, причем $M[X_{i}] = 0$, a $\sigma (X_{i}) = 1$. Тогда сумма квадратов этих величин

\begin{displaymath}
\chi ^2 = \sum\limits_{i = 1}^n {X_i ^2}
\end{displaymath}

распределена по закону $\chi^{2}$ с $k=n - 1$ степенями свободы. С увеличением числа степеней свободы распределение медленно приближается к нормальному.

**1. Проверка гипотезы о расхождении между эмпирическими (экспериментальными) частотами $\mbox{\boldmath$n_i$}$ и теоретическими (контрольными) частотами $\mbox{\boldmath$n'_i$}$ .**

**Алгоритм применения критерия $\chi^{2}$ Пирсона для сопоставления эмпирического и теоретического (другого эмпирического) распределений одного признака**

1. Занести в таблицу наименование разрядов и эмпирические частоты (данные по экспериментальной группе).
2. Во 2-й столбец записать теоретические частоты (данные по контрольной группе).
3. Проверить равенство сумм частот $\sum {n_i } = \sum {{n}'_i } $ (или их уравнять).
4. Подсчитать разности между эмпирическими и теоретическими частотами (экспериментальной и контрольной группами) по каждой строке и записать их в 3-й столбец.
5. Возвести в квадрат полученные разности и записать их в 4-й столбец.
6. Разделить полученные квадраты разностей на теоретические частоты (данные по контрольной группе) и записать в 5-й столбец.
7. Просуммировать значения 5-го столбца, обозначив ее $\chi_{эмп}^2 $
8. Определить по таблице критическое значение для соответствующего уровня значимости $\alpha $ и данного числа степеней свободы $r = m - 1$ ($m$ - количество разрядов признака, т.е. строк в таблице).
9. Если $\chi_{эмп}^2 \ge \chi_{кр}^2 $, то расхождения между распределениями существенны на данном уровне значимости.

**Пример 186.** При изучении творческой активности студентов были получены результаты для экспериментальных и контрольных групп. Определите, являются ли значимыми результаты предложенного подхода.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень усвоения  материала | Частота эксп. группа.  $n_{i}$ | Частота контр. группы  $n_{i}$' | ( $n_{i} - n_{i}')^{2}$ | ( $n_{i} - n_{i}')^{2}_{ / ni{\rm '}}$ |
| Хороший | 154 | 120 | 1156 | 9,63 |
| Прибл. | 36 | 49 | 169 | 3,44 |
| Плохой | 15 | 36 | 441 | 12,25 |
| Сумма | 205 | 205 |  | 25,32 |

7. $\chi_{эмп}^2 = 25,32$.

8. $\chi_{кр}^2 = 9,21$ для $\alpha = 0,01$ и $r = 2$ и поскольку $\chi_{эмп}^2 = 25,32 > 9,21 = \chi_{кр}^2 $, то нулевая гипотеза опровергается на высоком уровне значимости. Это позволяет признать, что разница частот контрольного и экспериментального ряда является статистически достоверной.

**Пример 187.**В банке в течение двух дней проводилось исследование времени обслуживания клиентов, результаты которого следующие:

|  |  |  |  |
| --- | --- | --- | --- |
| Номер интервала | Время обслуживания (мин) | Число клиентов в 1-й день | Число клиентов во 2-й день |
| 1 | 4 - 6 | 2 | 3 |
| 2 | 6 - 8 | 3 | 4 |
| 3 | 8 - 10 | 7 | 9 |
| 4 | 10 - 12 | 12 | 14 |
| 5 | 12 - 14 | 15 | 17 |
| 6 | 14 - 16 | 8 | 9 |
| 7 | 16 - 18 | 3 | 4 |

Можно ли считать одинаковыми среднее время обслуживания клиентов банка в первый и второй дни при $\alpha =0,05$?

Решение. Вычислим

\begin{displaymath}
\chi_{эмп}^2 = \sum\limits_{i = 1}^7 {{\displaystyle (n_i - n_i ')^2\over\displaystyle n_i '}}
=
\end{displaymath}

\begin{displaymath}
= {\displaystyle (2 - 3)^2\over\displaystyle 3} + {\display...
... +
{\displaystyle (3 - 4)^2\over\displaystyle 4} \approx 1,91.
\end{displaymath}

По таблице (прил. 5) критических точек распределения $\chi^2$ по заданному уровню значимости $\alpha =0,05$ и числу степеней свободы $k = 7-1$находим  критическую точку $\chi^2_{кр} (0,05;7) = 14,1$.

Поскольку $\chi_{эмп}^2 = 1,91 < 14,1 = \chi_{кр}^2$, то нет оснований отвергать нулевую гипотезу об одинаковом времени обслуживания клиентов банка в разные дни.

**2. Проверка гипотезы о нормальном распределении генеральной совокупности.** Для того, чтобы свести задачу к указанному в предыдущем пункте алгоритму, вычислим выборочные среднюю $х^{\ast }$ и среднее квадратическое отклонение $\sigma^{\ast}$, а затем - теоретические частоты $n_i' = {\displaystyle nh\over\displaystyle \sigma^\ast} \cdot \varphi(u_i)$, где $n$ - объем выборки, $h$ - шаг (разность между двумя соседними вариантами), $u_i = {\displaystyle x_i -
x^\ast \over\displaystyle \sigma ^\ast }$, $\varphi (u) = {\displaystyle 1\over\displaystyle \sqrt {2\pi} } \cdot e^{{
- u^2} \mathord{\left/ {\vphantom {{- u^2}2}} \right.
\kern-\nulldelimiterspace}2}$(приложение 1)

**Пример 188.**В результате выборочного обследования стажа работы профессорско-преподавательского состава получены следующие данные:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Стаж работы (лет) | 0 - 4 | 4 - 8 | 8 - 12 | 12 - 16 | 16 - 20 | 20 - 24 | 24 - 28 | 28 - 32 |
| Число  преподавателей | 3 | 8 | 25 | 40 | 46 | 31 | 6 | 2 |

Выясните, является ли распределение стажа работы нормальным при уровне значимости $\alpha = 0,01$.

Решение. Перейдем от заданного интервального распределения к распределению равноотстоящих вариант и вычислим выборочную среднюю и выборочное среднее квадратическое отклонение.

$n = 161, х^\ast = {\displaystyle 1\over\displaystyle 161}(2 \cdot 3 + 6 \cdot 8...
...+ 14
\cdot 40 + 18 \cdot 46 + 22 \cdot 31 + 26 \cdot 6 + 30 \cdot 2) \approx 16$,

\begin{displaymath}
\sigma ^\ast = \sqrt {D_х^\ast } \approx 5,43,{\rm а}
\quad
...
...yle \sigma ^\ast }\varphi (u_i ) = 118,6 \cdot \varphi
(u_i ).
\end{displaymath}

Составим расчетную таблицу, заполняя ее последовательно по столбцам слева направо

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| $i$ | $x_{i}$ | $u_i = (x_i - x^\ast ) / \sigma ^\ast $ | $\phi (u_{i})$ | $n_{i}'=75,8\phi (u_{i})$ | $n_{i}$ | $(n_{i} - n_{i}')^{2}$ | $(n_{i} - n_{i}')^{2}/ n_{i}'$ |
| 1 | 2 | -2,58 | 0,014 | 1,66 | 3 | 1,8 | 1,08 |
| 2 | 6 | -1,84 | 0,073 | 8,66 | 8 | 0,44 | 0,05 |
| 3 | 10 | -1,1 | 0,218 | 25,85 | 25 | 0,72 | 0,033 |
| 4 | 14 | -0,37 | 0,373 | 44,24 | 40 | 17,98 | 0,41 |
| 5 | 18 | 0,37 | 0,373 | 44,24 | 46 | 3,1 | 0,07 |
| 6 | 22 | 1,1 | 0,218 | 25,85 | 31 | 26,52 | 1,03 |
| 7 | 26 | 1,84 | 0,073 | 8,66 | 6 | 7,08 | 0,82 |
| 8 | 30 | 2,58 | 0,014 | 1,66 | 2 | 0,12 | 0,07 |
| $\sum $ |  |  |  |  | 161 |  | $\chi_{эмп}^2 = 3,56$ |

По таблице критических точек распределения (прил. 5) по уровню значимости $\alpha = 0,01$ и числу степеней свободы $k =s - 3 = 8 - 3 = 5$ находим критическую точку правосторонней критической области

\begin{displaymath}
\chi_{кр}^2 (0,01;5) = 15,1
\end{displaymath}

Так как $\chi_{эмп}^2 = 3,56 < 15,1 = \chi_{кр}^2 $, то гипотезу о нормальном распределении генеральной совокупности принимаем.

Пусть $Z$ - нормальная случайная величина, причем $M[Z] = 0$, $\sigma(Z) = 1$, a $V$ - независимая от $Z$ величина, которая распределена по закону $\chi^{2}$ с $k$ степенями свободы. Тогда величина

\begin{displaymath}
T = {\displaystyle Z\over\displaystyle \sqrt {V \mathord{\left/ {\vphantom {V k}} \right.
\kern-\nulldelimiterspace} k} }
\end{displaymath}

имеет распределение, которое называют $t$-pаспределением или распределением Стьюдента (псевдоним английского статистика В. Госсета), с $k$ степенями свободы. С возрастанием числа степеней свободы распределение Стьюдента быстро приближается к нормальному.

**Алгоритм применения $t$-критерия Стьюдента для сравнения оценки средних величин двух выборок**

1. Записать вариационный ряд результатов $Х$ экспериментальной группы.
2. Записать вариационный ряд результатов $Y$ контрольной группы.
3. Найти выборочные средние двух выборок $x^{\ast }$ и $y^{\ast}$.
4. Найти выборочные дисперсии $S_x^2 $ и $S_y^2 $.
5. Вычислить эмпирическое значение критической статистики

\begin{displaymath}
t_{эмп} = {\displaystyle \left\vert {x^\ast - y^\ast } \righ...
... {{\displaystyle n_1 \cdot
n_2 \over\displaystyle n_1 + n_2 }}
\end{displaymath}

Определить по таблице критическое значение $t_{кр} (\alpha ,n_1 + n_2
- 2)$ для соответствующего уровня значимости $\alpha $ и данного числа степеней свободы $r = n_1 + n_2 - 2$.

Если $t_{эмп} \ge t_{кр} $, то различия между средними значениями экспериментальной и контрольной групп существенны на данном уровне значимости.

**3. Сравнение двух средних нормальных генеральных совокупностей.** Для того, чтобы при заданном уровне значимости $\alpha $ проверить нулевую гипотезу $Н_{0} : М[Х] = М[Y]$ с неизвестными, но одинаковыми дисперсиями (в случае независимых малых выборок) при конкурирующей гипотезе $М[Х] \ne М[Y]$, надо вычислить наблюдаемое значение критерия:

\begin{displaymath}
t_{эмп} = {\displaystyle \left\vert {х^\ast - у^\ast } \righ...
... \sqrt {{\displaystyle nm(n + m - 2)\over\displaystyle n + m}}
\end{displaymath}

и по таблице критических точек распределения Стьюдента по заданному уровню значимости $\alpha $ и числу степеней свободы $k = n + m- 2$ найти критическую точку $t_{кр}(\alpha; k)$.

**Пример 189.**Для экспериментальной и контрольной групп в ходе   
5-тилетнего эксперимента чтения одинаковых лекционных курсов и проведения практических занятий при корреляции по годам средней успеваемости групп на начало эксперимента получено следующее варьирование разницы средних проявлений квазиисследовательской творческой деятельности студентов по годам для дисциплины теория вероятностей (с множителем 100):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| эксперимент | 5 | 6 | 7 | 6 | 7 | $\overline d_{э} = 6,2$ |
| контроль | 1 | 2 | 2 | 2 | 1 | $\overline d _{к} = 1,6$ |

Используя $t$-критерий Стьюдента на 1%-м уровне значимости, докажите достоверность различия между средними величинами опыта и контроля.

Решение. В нашем случае $х^{\ast } = 6,2$ и $y^{\ast } = 1,6$. Найдем исправленные дисперсии

\begin{displaymath}
S_х^2 = {\displaystyle 1\over\displaystyle 4}\left[ {(5 - 6,...
...2
\cdot 2} \right] \approx 0,7{\rm и}
\quad
S_у^2 \approx 0,3.
\end{displaymath}

Тогда $t_{эмп} = {\displaystyle 6,2 - 1,6\over\displaystyle \sqrt {4 \cdot 0,7 + 4 \cd...
...displaystyle 5 \cdot 5 \cdot (5 + 5 - 2)\over\displaystyle 5 + 5}} \approx 10,3$,

а по приложению 4 находим $t_{кр}(0,01;8) = 3,36$.

Поскольку $\left\vert {t_{эмп}} \right\vert = 10,3 > 3,36 = t_{кр} $, то нулевую гипотезу о незначительности различия между средними величинами опыта и контроля отвергают.

**4. Проверка гипотезы о значимости выборочного коэффициента корреляции.**Для того, чтобы при данном уровне значимости $\alpha $проверить нулевую  гипотезу $Н_{0}: r = 0$ o равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $H_{1}: r \ne 0 $, надо вычислить наблюдаемое значение критерия:

\begin{displaymath}
t_{эмп} = {r^\ast \sqrt {n - 2} } \mathord{\left/ {\vphantom...
... \right.
\kern-\nulldelimiterspace} {\sqrt {1 - (r^\ast )^2} }
\end{displaymath}

и по таблице критических точек распределения Стьюдента по заданному уровню значимости $\alpha $ и числу степеней свободы $k = n - 2$ найти критическую точку $t_{кр}(\alpha; k)$.

**Пример 190.** Определите, существует ли корреляционная связь между рождаемостью и смертностью для городов Ярославской области, используя информационно-статистические материалы по итогам 2002 года в сравнении с предыдущим.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Города | Ярославль | Рыбинск | Переславль | Ростов | Тутаев | Углич |
| Рождаемость  Х | 104,9 | 102,7 | 114,8 | 124,1 | 104,0 | 119,3 |
| Смертность  Y | 103,4 | 99,5 | 103,0 | 102,5 | 114,0 | 105,9 |

\begin{displaymath}
х^\ast = {\displaystyle 1\over\displaystyle 6}(107,9 + 102,7...
...24,1 + 104 + 119,3) \approx
111,6;
\quad
y^\ast \approx 104,7.
\end{displaymath}

2). Вычислим выборочную ковариацию

$k(X,Y) = {\displaystyle 1\over\displaystyle 6}\left[ {(104,9 - 111,6) \cdot (10...
....... + (119,3 - 111,6) \cdot \left. {(105,9 - 104,7)} \right] \approx
- 7,01.3)$. Найдем выборочные дисперсии

\begin{displaymath}
D_х^\ast = {\displaystyle 1\over\displaystyle 6}\left[ {(104...
...111,6)^2} \right] \approx 67,94;
\quad
D_у^\ast \approx 20,73.
\end{displaymath}

4). Вычислим выборочный коэффициент корреляции

\begin{displaymath}
r(X,Y) = {\displaystyle k(X,Y)\over\displaystyle \sqrt {D_х^...
...over\displaystyle \sqrt {67,94 \cdot 20,73} } \approx - 0,187.
\end{displaymath}

5). Вычислим наблюдаемое значение критерия

\begin{displaymath}
t_{эмп} = {\displaystyle r\sqrt {n - 2} \over\displaystyle \...
...2\over\displaystyle \sqrt {1 - ( - 0,187)^2} } \approx - 0,38.
\end{displaymath}

6). Выберем уровень значимости $\alpha = 0,1$ и найдем критическую точку $t_{кр}(0,1; 4) = 2,13$.

7). Поскольку $\left\vert {t_{эмп}} \right\vert = 0,38 < 2,13 = t_{кр}$, то нет оснований отвергать нулевую гипотезу $Н_{0}: r = 0$.

**Вопросы для самоконтроля**

1. Назовите основные типы статистических критериев проверки гипотезы.
2. В каких случаях применяются критерии Стьюдента и Пирсона?
3. Как определяются степени свободы?
4. Какие выборки следует считать однородными?
5. Какие критерии однородности вы знаете? Каковы условия применимости этих критериев?
6. Почему в критерии $\chi^{2}$ - Пирсона не может быть недопустимо малых значений критической статистики?
7. Коэффициенты ранговой корреляции Спирмена, Кендалла и критические точки распределения Стьюдента.
8. Приведите примеры практических задач, когда необходима проверка гипотез о равенстве математических ожиданий, дисперсий.

**Задачи**

**I** 371. По выборке объема $n = 16$, извлеченной из нормальной генеральной совокупности, найдены выборочная средняя $х^{\ast }= 59,1$ и "исправленное" среднее квадратическое отклонение $S^{\ast } = 1,8$. Проверьте нулевую гипотезу $Н_{0} : а = а_{0} = 60$ при конкурирующей гипотезе $Н_{1} : а \ne 60$и уровне значимости 0,05.

372. Проектный контролируемый размер изделий, изготавливаемых станком-автоматом, $а = а_{0} = 25$мм. Измерения 20 случайно отобранных изделий дали следующие результаты:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Контролируемый размер | 24,8 | 24,9 | 25,0 | 25,1 | 25,3 |
| Частота(число изделий) | 2 | 3 | 4 | 6 | 5 |

Проверьте нулевую гипотезу $Н_{0} : а = а_{0} = 25$ при конкурирующей гипотезе $Н_{1} : а \ne 25$ и уровне значимости 0,05.

373. Физическая подготовка 10 спортсменов была проверена при поступлении на факультет физической культуры, а затем после первого семестра. Итоги проверки в баллах оказались следующими:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Число баллов при поступлении | 75 | 72 | 55 | 48 | 60 | 68 | 28 | 65 | 58 | 71 |
| Число баллов после ${\rm I}$ семестра | 80 | 85 | 51 | 52 | 60 | 62 | 35 | 83 | 61 | 87 |

Установите при уровне значимости 0,1, значимо или незначимо улучшилась физическая подготовка спортсменов, в предположении, что число баллов распределено нормально.

374. Известны следующие выборочные данные о посеве яровых за день (в процентах к плану) хозяйствами области: 3, 4, 7, 1, 3, 3, 1, 5, 6, 2, 4, 3, 4, 5, 2, 2, 4, 3, 5, 6, 4, 3, 2, 5, 4.

Составьте вариационный ряд, постройте полигон относительных частот и проверьте по критерию Пирсона, что распределение нормальное.

375. Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать $\sigma _0^2 = 0,1$. Взята проба из 25 случайно отобранных изделий, которая дала следующие результаты измерений:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| контролируемый размер | 3,0 | 3,5 | 3,8 | 4,4 | 4,5 |
| частота | 2 | 6 | 9 | 7 | 1 |

Проверьте, обеспечивает ли станок требуемую точность при уровне значимости 0,05.

376. Исследование в течение 10 дней производительности двух предприятий, выпускающих стиральные машины, дало следующие результаты:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1-е предприятие | 82 | 74 | 64 | 72 | 84 | 68 | 76 | 88 | 75 | 60 |
| 2-е предприятие | 52 | 63 | 72 | 64 | 48 | 70 | 78 | 68 | 75 | 54 |

Можно ли считать одинаковыми распределения производительности выпуска стиральных машин на обоих предприятиях при уровне значимости $\alpha =0,05$?

**II** 377. Во время медосбора были получены следующие выборочные данные о суточной прибавке меда на пасеках области:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Суточная прибавка (кг.) | 3 - 4 | 4 - 5 | 5 - 6 | 6 - 7 | 7 - 8 | 8 - 9 |
| Количество пчелосемей | 6 | 26 | 55 | 45 | 15 | 3 |

Подчиняется ли суточная прибавка меда нормальному закону?

378. Найдите выборочный коэффициент ранговой корреляции для измерениями двумя приборами шести деталей (в сотых долях миллиметра):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| $x_{i}$ | 11 | 10 | 12 | 9 | 13 | 8 |
| $y_{i}$ | 12 | 9 | 13 | 8 | 14 | 7 |

Установите, при каком уровне значимости $\alpha $ ранговая корреляционная связь будет значимой.

**III** 379. Результаты выборочного обследования коров стада одного сельхозпредприятия по годовому надою приведены в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Годовой надой (тыс кг.) | 2,7-2,9 | 2,9-3,1 | 3,1-3,3 | 3,3-3,5 | 3,5-3,7 | 3,7-3,9 | 3,9-4,1 | 4,1-4,3 |
| Количество коров | 5 | 16 | 33 | 43 | 37 | 20 | 7 | 2 |

Выясните, является ли распределение годового надоя нормальным. Найдите процент надоев в среднем по хозяйству не менее 3500 кг и процент коров, у которых средний годовой надой отличается от среднего годового надоя по хозяйству меньше чем на 500 кг.

380. По выборке объема $n = 100$, известной из двумерной нормальной генеральной совокупности $(Х,Y)$, составлена корреляционная таблица:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y  Х | 10 | 15 | 20 | 25 | 30 | 35 | $n_{х}$ |
| 25 | 5 | 1 | - | - | - | - | 6 |
| 35 | - | 6 | 2 | - | - | - | 8 |
| 45 | - | - | 5 | 40 | 5 | - | 50 |
| 55 | - | - | 2 | 8 | 7 | - | 17 |
| 65 | - | - | - | 4 | 7 | 8 | 19 |
| $n_{у}$ | 5 | 7 | 9 | 52 | 19 | 8 | $n = 100$ |

Найдите выборочный коэффициент корреляции и проверьте нулевую гипотезу о равенстве генерального коэффициента корреляции нулю при конкурирующей гипотезе $Н_{1}: r \ne 0$ и при уровне значимости 0,05.

|  |  |  |
| --- | --- | --- |
|  |  |  |

**Далее:** [§39. Математическая статистика в](http://cito-web.yspu.org/link1/metod/theory/node45.html) **Вверх:** [Глава V. Математическая статистика](http://cito-web.yspu.org/link1/metod/theory/node37.html) **Назад:** [§37. Статистические гипотезы](http://cito-web.yspu.org/link1/metod/theory/node43.html)

[*ЯГПУ, Центр информационных технологий обучения*](http://www.yspu.yar.ru/cito)*2006-03-04*