Министерство образования и науки
Российской Федерации
Федеральное агентство по образованию

Алтайский государственный университет

Математический факультет

Кафедра математического анализа

# Распределения, связанные с нормальным

Барнаул, 2007

Распределения, связанные с нормальным (Методические указания для студентов математического и физического факультета). Составитель Дронов С.В. (к.ф.-м.н. доцент) .- Барнаул, 2007.- 44 с.

Собраны основные свойства нормального распределения и связанных с ним – Г-распределений, распределений Стьюдента и Фишера. Даны необходимые таблицы, а также описаны способы решения задач, входящих в семестровое расчетное задание по математической статистике.

Рецензент – Саженков А.Н. (к.ф.-м.н., доцент)

Методические указания также выставлены для свободного доступа на интернет-сайте математического факультета по адресу www.math.asu.ru.

# Оглавление

# 1 Нормальное распределение на прямой

Рассмотрим при некоторых значениях $a$ и $\sigma > 0$ следующую функцию

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \tag{1}$$

**Лемма 1.** *При $a = 0$ , $\sigma = 1$ функция (1) задает плотность некоторого вероятностного распределения на прямой.*

**Доказательство**. Для проверки утверждения леммы достаточно показать, что

$$I = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} dx = 1.$$

Запишем следующее очевидное равенство

$$I^2 = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \exp\left\{-\frac{x^2+y^2}{2}\right\} dx\ dy.$$

Перейдем к полярным координатам:

$$I^2 = \frac{1}{2\pi} \int\limits_0^{2\pi} d\varphi \int\limits_0^{\infty} r \exp\left\{-\frac{r^2}{2}\right\} dr \ = \ \int\limits_0^{\infty} \exp\left\{-\frac{r^2}{2}\right\} d\left(\frac{r^2}{2}\right) = 1.$$

Поскольку $I > 0$, то лемма доказана.

**Лемма 2.** *При произвольных $a \in \mathbf{R}$ и $\sigma > 0$ функция (1) есть плотность вероятностного распределения.*

Это утверждение немедленно следует из леммы 1, если в $I$ мы сделаем замену $x = \frac{y-a}{\sigma}$.

Распределение с плотностью (1) называется *нормальным* с параметрами $a, \sigma^2$. Если $\xi$ имеет такое распределение, то условимся писать $\xi \models N(a, \sigma^2)$. Вообще, знак $\models$ будем использовать вместо слов "имеет распределение".

Распределение $N(0,1)$ будем называть *стандартным нормальным*, а обозначение $\Phi$ закрепим за его функцией:

$$\Phi(x) \ = \ \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\{-y^2/2\} dy.$$

**Теорема 1.** *Пусть* $\xi \models N(a, \sigma^2)$. *Тогда* $\eta = \alpha\xi + \beta \quad (\alpha \neq 0)$ *имеет распределение* $N(\alpha a + \beta, \alpha^2\sigma^2)$.

**Доказательство**. Действительно,

$$\mathbf{P}(\eta < x) \ = \ \mathbf{P}(\alpha\xi < x - \beta),$$

и, разрешая это неравенство относительно $\xi$ (конечно, придется отдельно рассмотреть случаи положительного и отрицательного $\alpha$), можно в полученном интеграле сделать замену $y \ = \ \alpha x + \beta$. Получим

$$\mathbf{P}(\eta < x) \ = \ \frac{1}{\mid \alpha \mid \sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-\frac{(y - \alpha a - \beta)^2}{2\sigma^2\alpha^2}\right\} \, dy.$$

Поскольку последнее равенство выполнено при произвольном $x$, то плотность $\eta$ имеет вид

$$\frac{1}{\mid \alpha \mid \sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \alpha a - \beta)^2}{2\sigma^2\alpha^2}\right\},$$

откуда $\eta \ \models \ N(\alpha a + \beta, \alpha^2\sigma^2)$.

**Следствие 1.** *Пусть* $\xi \ \models \ N(a, \sigma^2)$. *Тогда* $\frac{\xi - a}{\sigma} \ \models \ N(0,1)$.

Для доказательства достаточно положить в теореме $\alpha \ = \ \frac{1}{\sigma}$, $\beta = - \ \frac{a}{\sigma}$.

## 2 Характеристическая функция

**Теорема 2.** *Пусть $\xi$ имеет стандартное нормальное распределение. Тогда ее характеристическая функция имеет вид*

$$\varphi(t) \;=\; \exp\left\{-\,\frac{t^2}{2}\right\}.$$

**Доказательство**. Поскольку в равенстве

$$\mathbf{M}\mid\xi\mid \quad = \quad \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\mid x\mid\exp\left\{-\,\frac{x^2}{2}\right\}\,dx$$

интеграл справа, очевидно, сходится, то $\varphi(t)$ - дифференцируемая функция, причем, интегрируя по частям, получим

$$\varphi'(t) \;=\; \tfrac{i}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x\exp\{itx\}\exp\{-\,\tfrac{x^2}{2}\}\;dx \;=$$
$$=\;-\;\tfrac{i}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\{itx\}\;d\exp\{-\,\tfrac{x^2}{2}\}\;=$$
$$=\;-\tfrac{i}{\sqrt{2\pi}}\exp\{itx\}\exp\{-x^2/2\}|_{-\infty}^{\infty}-t\varphi(t),$$

или $\varphi'(t) \;=\; -t\varphi(t)$, откуда $\varphi(t) \;=\; C\exp\{-\frac{t^2}{2}\}$. Осталось заметить, что $C=1$, поскольку $\varphi(0)=1$. Теорема доказана.

**Следствие 2.** *Распределение $N(0,1)$ является симметричным, т.е. если $\xi\models\; N(0,1)$, то $(-\,\xi)\models N(0,1)$.*

Действительно, т.к. в силу теоремы 2 $\varphi_{-\xi}(t) \;=\; \varphi(-t) \;=\; \varphi(t)$, то распределения $-\xi$ и $\xi$ совпадают. Заметим также, что этот факт легко вывести и из теоремы 1, если взять в ее условии $\alpha=-1,\;\beta=0$.

**Следствие 3.** *Функция стандартного нормального распределения обладает следующими свойствами:*

- $\Phi(x) \ + \ \Phi(-x) \ = \ 1$ *для произвольного* $x$*;*
- $\Phi(0) \ = \ \frac{1}{2}$.

**Доказательство**. Согласно следствию 2,

$$\Phi(-x) \ = \ \mathbf{P}(-\xi > x) \ = \ \mathbf{P}(\xi > x) \ = \ 1 - \Phi(x),$$

что совпадает с первым утверждением. Если теперь в этом утверждении выберем $x = 0$, то немедленно получим и второе.

**Следствие 4.** *Характеристическая функция* $N(a, \sigma^2)$ *равна*

$$\varphi(t) \ = \ \exp\left\{ita \ - \ \frac{t^2\sigma^2}{2}\right\}.$$

Действительно, пусть $\xi \ \models \ N(a, \sigma^2)$. Согласно следствию 1, $\eta \ \equiv \ \frac{\xi - a}{\sigma} \ \models N(0,1)$. Таким образом, $\xi \ = \ \sigma\eta + a$, и из теоремы 2 получаем

$$\varphi_\xi(t) \ = \ e^{ita}\varphi_\eta(\sigma t) \ = \ \exp\left\{ita \ - \ \frac{t^2\sigma^2}{2}\right\}.$$

**Следствие 5.** *Пусть случайные величины* $\xi_1, ..., \xi_n$ *независимы,* $\xi_j \models N(a_j, \sigma_j^2)$*,* $j = 1, ..., n$*. Тогда*

$$S_n \ \equiv \ \sum_{j=1}^{n} \xi_j \ \models \ N(\sum_{j=1}^{n} a_j, \ \sum_{j=1}^{n} \sigma_j^2).$$

Действительно, из следствия 4 и независимости вытекает

$$\varphi_{S_n}(t) \ = \ \prod_{j=1}^{n} \varphi_{\xi_j}(t) \ = \ \exp\{it\sum_{j=1}^{n} a_j \ - \ \frac{t^2}{2}\sum_{j=1}^{n} \sigma_j^2\}.$$

Утверждение следствия теперь получается из теоремы единственности для характеристических функций.

**Следствие 6.** *Пусть* $\xi \models N(0,1)$. *Тогда при нечетных* $n$ $\mathbf{M}\xi^n = 0$, *а при четных* $\mathbf{M}\xi^n = (n-1)!!$.

**Доказательство**. Поскольку

$$\varphi_\xi(t) = \exp\left\{-\frac{t^2}{2}\right\} = 1 + \sum_{k=1}^{\infty}(-1)^k \frac{t^{2k}}{k!2^k},$$

и, с другой стороны, т.к. $\varphi$ - характеристическая, то

$$\varphi_\xi(t) = 1 + \sum_{n=1}^{\infty} (it)^n \frac{\mathbf{M}\xi^n}{n!}.$$

Отсюда заключаем, что при $n = 2k$

$$\frac{\mathbf{M}\xi^{2k}}{(2k)!} = \frac{1}{k!2^k} = \frac{1}{(2k)!!},$$

а при $n = 2k-1 \quad \mathbf{M}\xi^n = 0$. Окончательно,

$$\mathbf{M}\xi^{2k} = \frac{(2k)!}{(2k)!!} = (2k-1)!!.$$

**Следствие 7.** *Если* $\xi \models N(a,\sigma^2)$, *то* $\mathbf{M}\xi = a$, $\mathbf{D}\xi = \sigma^2$.

Пусть $\eta \models N(0,1)$. По следствию 6, $\mathbf{M}\eta = 0, \mathbf{D}\eta = \mathbf{M}\eta^2 = 1$. Поскольку $\xi$ можно представить в виде $\xi = \sigma\eta + a$, то

$$\mathbf{M}\xi = \sigma\mathbf{M}\eta + a = a, \quad \mathbf{D}\xi = \sigma^2\mathbf{D}\eta = \sigma^2.$$

Заметим, что утверждение следствия может быть получено из формулы (1) интегрированием.

## 3 Таблицы нормального распределения

Пусть $\xi \models N(a,\sigma^2)$. Объединяя утверждения следствия 1 с тривиальным равенством

$$\mathbf{P}(\xi < x) = \mathbf{P}(\eta < \frac{x-a}{\sigma}) = \Phi\left(\frac{x-a}{\sigma}\right),$$

видим, что для вычисления вероятностей, связанных с произвольным нормальным распределением, достаточно иметь таблицы $\Phi(x)$. Если же мы к тому же будем иметь ввиду равенства следствия 3, то становится ясно, что можно иметь таблицы этой функции только для $x$, имеющих определенный знак, например, для $x < 0$, как и сделано в [1], откуда заимствована приводимая таблица.

Для положительных аргументов пользуйтесь формулой

$$\Phi(x) \; = \; 1 - \Phi(-x).$$

Еще пара значений для справки

$$\Phi(-4) = 0.000031671, \;\; \Phi(-5) = 0.00000029665.$$

Полезно также помнить о том, что табулированная функция монотонно возрастает.

Учитывая малую информативность $\Phi(x)$ при больших по модулю $x$, иногда табулируют функцию

$$\Phi_{gm}(x) \quad = \quad \frac{1}{\sqrt{2\pi}} \int\limits_0^x \exp\left\{-\frac{y^2}{2}\right\} \, dy$$

(см., например, [2]). Получим связь $\Phi$ и $\Phi_{gm}$. Поскольку $\Phi(0) = 0.5$, то при $x > 0$ имеем

$$\Phi(x) \; = \; \Phi(0) \; + \; \frac{1}{\sqrt{2\pi}} \int\limits_0^x \exp\left\{-\frac{y^2}{2}\right\} dy \; = \; \frac{1}{2} \; + \; \Phi_{gm}(x). \quad (2)$$

Если же $x < 0$, то

$$\Phi(x) \; = \; \Phi(0) \; - \; \frac{1}{\sqrt{2\pi}} \int\limits_x^0 \exp\{-\frac{y^2}{2}\} dy \; = \; \frac{1}{2} \; - \; \Phi_{gm}(-x). \quad (3)$$

## Т А Б Л И Ц А 1

Функция стандартного нормального распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\{-\frac{y^2}{2}\}\, dy$$

| $x$ | 0 | 2 | 4 | 6 | 8 |
|---|---|---|---|---|---|
| -0.0 | 0.5000 | 0.4920 | 0.4840 | 0.4761 | 0.4681 |
| -0.1 | 0.4602 | 0.4522 | 0.4443 | 0.4364 | 0.4286 |
| -0.2 | 0.4207 | 0.4129 | 0.4052 | 0.3974 | 0.3897 |
| -0.3 | 0.3821 | 0.3745 | 0.3669 | 0.3594 | 0.3520 |
| -0.4 | 0.3446 | 0.3372 | 0.3300 | 0.3228 | 0.3156 |
| -0.5 | 0.3085 | 0.3015 | 0.2946 | 0.2877 | 0.2810 |
| -0.6 | 0.2743 | 0.2676 | 0.2611 | 0.2546 | 0.2483 |
| -0.7 | 0.2420 | 0.2358 | 0.2297 | 0.2236 | 0.2177 |
| -0.8 | 0.2119 | 0.2061 | 0.2005 | 0.1949 | 0.1894 |
| -0.9 | 0.1841 | 0.1788 | 0.1736 | 0.1685 | 0.1635 |
| -1.0 | 0.1587 | 0.1539 | 0.1492 | 0.1446 | 0.1401 |
| -1.1 | 0.1357 | 0.1314 | 0.1271 | 0.1230 | 0.1190 |
| -1.2 | 0.1151 | 0.1112 | 0.1075 | 0.1038 | 0.1003 |
| -1.3 | 0.0968 | 0.0934 | 0.0901 | 0.0869 | 0.0838 |
| -1.4 | 0.0808 | 0.0778 | 0.0749 | 0.0721 | 0.0694 |
| -1.5 | 0.0668 | 0.0643 | 0.0618 | 0.0594 | 0.0571 |
| -1.6 | 0.0548 | 0.0526 | 0.0505 | 0.0485 | 0.0465 |
| -1.7 | 0.0446 | 0.0427 | 0.0409 | 0.0392 | 0.0375 |
| -1.8 | 0.0359 | 0.0344 | 0.0329 | 0.0314 | 0.0301 |
| -1.9 | 0.0288 | 0.0274 | 0.0262 | 0.0250 | 0.0239 |
| -2.0 | 0.0228 | 0.0217 | 0.0207 | 0.0197 | 0.0188 |
| -2.1 | 0.0179 | 0.0170 | 0.0162 | 0.0154 | 0.0146 |
| -2.2 | 0.0139 | 0.0132 | 0.0125 | 0.0119 | 0.0113 |
| -2.3 | 0.0107 | 0.0102 | 0.0096 | 0.0091 | 0.0087 |
| -2.4 | 0.0082 | 0.0078 | 0.0073 | 0.0069 | 0.0066 |

| $x$ | 0 | 2 | 4 | 6 | 8 |
|---|---|---|---|---|---|
| -2.5 | 0.0062 | 0.0059 | 0.0055 | 0.0052 | 0.0049 |
| -2.6 | 0.0047 | 0.0044 | 0.0041 | 0.0039 | 0.0037 |
| -2.7 | 0.0035 | 0.0033 | 0.0031 | 0.0029 | 0.0027 |
| -2.8 | 0.0026 | 0.0024 | 0.0023 | 0.0021 | 0.0020 |
| -2.9 | 0.0019 | 0.0018 | 0.0016 | 0.0015 | 0.0014 |
| -3.0 | 0.0013 | 0.0013 | 0.0012 | 0.0011 | 0.0010 |
| -3.1 | 0.0010 | 0.0009 | 0.0008 | 0.0008 | 0.0007 |
| -3.2 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0005 |
| -3.3 | 0.0005 | | | | |
| -3.4 | 0.0003 | | | | |
| -3.5 | 0.0002 | | | | |
| -3.6 | 0.0002 | | | | |
| -3.7 | 0.0001 | | | | |
| -3.8 | 0.0001 | | | | |
| -3.9 | 0.0000 | | | | |

Иногда для $\eta \models N(0,1)$ полезной оказывается формула

$$\mathbf{P}(\mid \eta \mid < x) \;=\; \frac{1}{\sqrt{2\pi}} \int\limits_{-x}^{x} \exp\{-\frac{y^2}{2}\}\, dy \;=\; 2\Phi_{gm}(x). \qquad (4)$$

Отметим также, что

$$\Phi(x) \;-\; \Phi(-x) \;= 1 - 2\Phi(-x) \;=\; \mathbf{P}(\mid \eta \mid < x). \qquad (5)$$

Для удобства приведем таблицу наиболее полезных *критических точек* $\Phi$ *уровня* $\varepsilon$, то есть таких $t_\varepsilon$, что $\mathbf{P}(\eta > t_\varepsilon) = \varepsilon$. При этом

$$t_{1-\varepsilon} \;=\; -t_\varepsilon; \qquad t_{double,\varepsilon} \;=\; t_{1-\frac{\varepsilon}{2}},$$

где $t_{double,\varepsilon}$ выбрано так, что

$$\mathbf{P}(\mid \eta \mid > t_{double,\varepsilon}) \;=\; \varepsilon$$

и называется *двусторонней критической точкой уровня* $\varepsilon$ *распределения* $\eta$.

Т А Б Л И Ц А 2

Критические точки стандартного нормального распределения

| $\varepsilon$ | 0.0025 | 0.005 | 0.01 | 0.025 | 0.05 | 0.1 |
|---|---|---|---|---|---|---|
| $t_\varepsilon$ | -2.89 | -2.57 | -2.32 | -1.96 | -1.64 | -1.26 |

# 4 Гамма-распределения. Распределение $\chi^2$

Тесно связано с нормальным так называемое $\Gamma_{\alpha,\lambda}$-*распределение* с плотностью

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-x} \qquad (x \geq 0). \tag{6}$$

Здесь $\alpha, \lambda$ – положительные параметры, а

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx-$$

интеграл Эйлера. Из курса математического анализа известно, что при $\lambda > 0$ $\Gamma(\lambda+1) = \lambda\Gamma(\lambda)$. Выполнено также $\Gamma(n) = (n-1)!$ и $\Gamma(\frac{1}{2}) = \frac{1}{\sqrt{\pi}}$.

Найдем характеристическую функцию распределения $\Gamma_{\alpha,\lambda}$. По определению,

$$\varphi(t) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty x^{\lambda-1} e^{itx-\alpha x} dx.$$

Поскольку

$$| x^\lambda e^{itx-\alpha x} | = | x |^\lambda e^{-\alpha x},$$

и интеграл от правой части последнего равенства сходится, то функция $\varphi$ дифференцируема, причем

$$\varphi'(t) \;=\; -\frac{i\alpha^\lambda}{\Gamma(\lambda)}\int\limits_0^\infty x^\lambda e^{itx-\alpha x}dx \;=\; \frac{i\alpha^\lambda}{\Gamma(\lambda)}\int\limits_0^\infty x^\lambda d\left(\frac{e^{(it-\alpha)x}}{it-\alpha}\right).$$

Производя интегрирование по частям, имеем

$$\varphi'(t) \;=\; -\frac{i\lambda\alpha^\lambda}{(it-\alpha)\Gamma(\lambda)}\int\limits_0^\infty x^{\lambda-1}e^{itx-\alpha x}dx \;=\; -\frac{\lambda}{t+i\alpha}\varphi(t).$$

Решая это дифференциальное уравнение, получим

$$\varphi(t) \;=\; C(t+i\alpha)^{-\lambda}.$$

Т.к. $\varphi(0)=1$, то $C \;=\; (i\alpha)^\lambda$, и окончательно

$$\varphi(t) \;=\; \left(1-\frac{it}{\alpha}\right)^{-\lambda}. \tag{7}$$

Пусть теперь $\xi_1,...,\xi_n \;\models\; N(0,1)$ и независимы. Покажем, что распределение случайной величины $\eta \;=\; \xi_1^2+...+\xi_n^2$, называемое $\chi^2$*-распределением с $n$ степенями свободы* (при этом пишут $\eta\models\chi_n^2$), есть частный случай Г-распределения.

**Теорема 3.**

$$\chi_n^2 \quad = \quad \Gamma_{1/2,\ n/2}.$$

**Доказательство**. Поскольку $\xi_1\models N(0,1)$, то при $x>0$ выполнено

$$\mathbf{P}(\xi_1^2<x) \;=\; \mathbf{P}(\mid\xi_1\mid<\sqrt{x}) \;=\; 2\Phi(\sqrt{x})-1.$$

Дифференцируя это равенство по $x$, убеждаемся, что плотность распределения $\xi_1^2$ есть плотность $\Gamma_{\frac{1}{2},\frac{n}{2}}$ (см. (6)). Поэтому, согласно (7),

$$\varphi_{\xi_1^2}(t) \;=\; (1-2it)^{-\frac{1}{2}}.$$

То же самое справедливо для каждого из $\xi_j^2$. Т.к. они независимы, то

$$\varphi_\eta(t) \;=\; \varphi_{\Sigma\xi_j^2}(t) \;=\; (\varphi_{\xi_1^2}(t))^n \;=\; (1-2it)^{-\frac{n}{2}}.$$

Теорема доказана.

Ниже приводится таблица критических точек $\chi^2$, заимствованная из [3].

## 5 Распределения Стьюдента и Фишера

Пусть $\xi \models N(0,1), \quad \eta \models \chi_n^2$ и они независимы. *Распределением Стьюдента с $n$ степенями свободы* $T_n$ называется распределение случайной величины $\frac{\xi\sqrt{n}}{\sqrt{\eta}}$. В принципе, поскольку плотности распределений $\xi$ и $\eta$ нам известны, то нетрудно найти и плотность $T_n$. Она равна

$$f(x) \quad = \quad \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1 \;+\; \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Можно также заметить несколько полезных свойств этого распределения. Во-первых, оно симметрично. Действительно, если $\zeta \models T_n$, то, воспользовавшись симметричностью $N(0,1)$, имеем

$$-\zeta \;=\; \frac{-\xi\sqrt{n}}{\sqrt{\eta}} \;\stackrel{d}{=}\; \frac{\xi\sqrt{n}}{\sqrt{\eta}} \;=\; \zeta,$$

где значок $\stackrel{d}{=}$ означает совпадение распределений. Отсюда, как и для нормального распределения, легко вывести, что для произвольного $x$

$$T_n(x) \;+\; T_n(-x) \;=\; 1 \qquad (T_n(x) = \mathbf{P}(\zeta < x)).$$

## Т А Б Л И Ц А 3

Критические точки $t_\varepsilon$ распределений $\chi^2_n$:
$\mathbf{P}(\chi^2_n > t_\varepsilon) = \varepsilon$

| $n \downarrow \varepsilon \rightarrow$ | 0.995 | 0.99 | 0.95 | 0.05 | 0.01 | 0.005 |
|---|---|---|---|---|---|---|
| 2 | 0.01 | 0.02 | 0.10 | 5.99 | 9.21 | 10.60 |
| 3 | 0.07 | 0.11 | 0.35 | 7.81 | 11.34 | 12.84 |
| 4 | 0.21 | 0.30 | 0.71 | 9.49 | 13.28 | 14.86 |
| 5 | 0.41 | 0.55 | 1.14 | 11.07 | 15.09 | 16.75 |
| 6 | 0.68 | 0.87 | 1.63 | 12.59 | 16.81 | 18.56 |
| 7 | 0.99 | 1.24 | 2.17 | 14.07 | 18.47 | 20.28 |
| 8 | 1.34 | 1.65 | 2.73 | 15.51 | 20.09 | 21.96 |
| 9 | 1.74 | 2.09 | 3.33 | 16.92 | 21.67 | 23.59 |
| 10 | 2.16 | 2.56 | 3.94 | 18.31 | 23.21 | 25.19 |

| $n \downarrow \varepsilon \rightarrow$ | 0.995 | 0.99 | 0.95 | 0.05 | 0.1 | 0.005 |
|---|---|---|---|---|---|---|
| 12 | 3.07 | 3.57 | 5.23 | 21.03 | 26.22 | 28.30 |
| 14 | 4.07 | 4.66 | 6.57 | 23.68 | 29.14 | 31.32 |
| 16 | 5.14 | 5.81 | 7.96 | 26.30 | 32.00 | 34.27 |
| 18 | 6.26 | 7.01 | 9.39 | 28.87 | 34.80 | 37.16 |
| 20 | 7.43 | 8.26 | 10.85 | 31.41 | 37.57 | 40.00 |
| 21 | 8.03 | 8.90 | 11.59 | 32.67 | 38.93 | 41.40 |
| 22 | 8.64 | 9.54 | 12.34 | 33.92 | 40.29 | 42.80 |
| 23 | 9.26 | 10.20 | 13.09 | 35.17 | 41.64 | 44.17 |
| 24 | 9.89 | 10.86 | 13.85 | 36.41 | 42.98 | 45.56 |
| 25 | 10.52 | 11.52 | 14.61 | 37.65 | 44.31 | 46.93 |
| 30 | 13.79 | 14.95 | 18.49 | 43.77 | 50.89 | 53.67 |
| 35 | 17.19 | 18.51 | 22.49 | 49.80 | 57.34 | 60.27 |
| 40 | 20.71 | 22.16 | 26.51 | 55.76 | 63.69 | 66.77 |
| 45 | 24.31 | 25.90 | 30.61 | 61.66 | 69.96 | 73.17 |
| 50 | 27.99 | 29.71 | 34.76 | 67.50 | 76.15 | 79.49 |

Во-вторых, поскольку

$$\frac{1}{n}\eta \ \stackrel{d}{=} \ \frac{1}{n}\sum_{j=1}^{n}\xi_j^2, \quad \xi_1, ..., \xi_n \ \models N(0,1)$$

и независимы, то, согласно закону больших чисел,

$$\frac{1}{n}\eta \ \stackrel{d}{\to} \ \mathbf{M}\xi_1^2 \ = \ 1 \qquad (n \to \infty),$$

следовательно,

$$\zeta \ \stackrel{d}{\to} \ \frac{\xi}{\sqrt{\mathbf{M}\xi_1^2}} \ = \ \xi \ \models \ N(0,1) \qquad (n \to \infty),$$

т.е. при большом числе степеней свободы распределение Стьюдента практически совпадает со стандартным нормальным.

В приложениях часто встречается *распределение Фишера* (или $F$-распределение) $F_{n,m}$. Этим распределением обладает случайная величина $\frac{m\eta}{n\xi}$, где $\xi \models \chi_n^2, \eta \models \chi_m^2$ и они независимы. Очевидно, что если $t_{\varepsilon,n,m}$ - критическая точка $F_{n,m}$ уровня $\varepsilon$, то

$$\varepsilon \ = \ \mathbf{P}(\tfrac{m\eta}{n\xi} > t_{\varepsilon,n,m}) \ = \ \mathbf{P}\left(\tfrac{n\xi}{m\eta} < \tfrac{1}{t_{\varepsilon,n,m}}\right) = \\ = \ 1 - \mathbf{P}\left(\tfrac{n\xi}{m\eta} > \tfrac{1}{t_{\varepsilon,n,m}}\right),$$

то есть $\frac{1}{t_{\varepsilon,n,m}}$ - критическая точка $F_{m,n}$ уровня $1-\varepsilon$.

Далее приведены таблицы критических точек распределений Стьюдента и Фишера, заимствованные из [3].

# 6 Многомерное нормальное распределение

*Многомерным нормальным распределением* называется распределение в $\mathbf{R}^n$ с плотностью

$$f(\vec{x}) \ = \ \frac{\sqrt{| \ A \ |}}{(\sqrt{2\pi})^n}\exp\{-\ \frac{1}{2} < A(\vec{x}-\vec{a}),(\vec{x}-\vec{a}) >\}, \qquad (8)$$

ТАБЛИЦА 4

Критические точки $t(\varepsilon)$ распределения Стьюдента $T_n$.
В таблице собраны односторонние критические точки: $\mathbf{P}(\zeta > t(\varepsilon)) = \varepsilon$.
Для нахождения $\tau(\varepsilon)\ :\ \mathbf{P}(|\ \zeta\ |> \tau(\varepsilon)) = \varepsilon$
следует положить $\tau(\varepsilon)\ =\ t(\frac{\varepsilon}{2})$.

| $n \downarrow\ \ \varepsilon \rightarrow$ | 0.05 | 0.005 | 0.0005 | $n \downarrow\ \ \varepsilon \rightarrow$ | 0.05 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|
| 1 | 5.31 | 63.66 | 636.62 | 21 | 2.08 | 2.93 | 3.82 |
| 2 | 2.92 | 9.92 | 31.60 | 22 | 2.07 | 2.82 | 3.79 |
| 3 | 2.35 | 5.84 | 12.92 | 23 | 2.07 | 2.81 | 3.77 |
| 4 | 2.13 | 4.60 | 8.61 | 24 | 2.06 | 2.80 | 3.75 |
| 5 | 2.02 | 4.03 | 6.87 | 25 | 2.06 | 2.79 | 3.73 |
| 10 | 2.23 | 3.17 | 4.59 | 30 | 2.04 | 2.75 | 3.65 |
| 12 | 2.18 | 3,06 | 4.33 | 36 | 2.03 | 2.72 | 3.58 |
| 15 | 2.13 | 2.95 | 4.07 | 40 | 2.02 | 2.70 | 3.55 |
| 17 | 2.11 | 2.90 | 3.96 | 46 | 2.01 | 2.69 | 3.51 |
| 20 | 2.09 | 2.85 | 3.85 | 50 | 2.01 | 2.68 | 3.50 |
| | | | | $\infty$ | 1.96 | 2.57 | 3.29 |

где $A$ – симметричная положительно определенная матрица, $|\ A\ |$– ее определитель, $\vec{a} \in \mathbf{R}^n$ – некоторый вектор, символом $<.,.>$ обозначено скалярное произведение.

В том случае, когда $\vec{a} = \vec{0}$, распределение называют *центрированным*. Если же у центрированного нормального распределения в (8) $A$ есть единичная $n \times n$ матрица, то такое распределение условимся называть *стандартным нормальным*. Аналогично одномерному случаю имеет место

**Лемма 3.** *Если $\vec{\xi} \in \mathbf{R}^n$ обладает плотностью (8), то распределение $\vec{\eta}\ =\ A^{\frac{1}{2}}(\vec{\xi} - \vec{a})$ является стандартным нормальным.*

**Доказательство**. Пусть $\vec{x}\ =\ (x_1, ..., x_n)$. Условимся писать $\vec{\xi} < \vec{x}$, если неравенства $\xi_1 < x_1,\ \xi_2 < x_2,\ ...\ \xi_n < x_n$ выполнены

Т А Б Л И Ц А 5

Распределение Фишера $F_{n,m}$

В таблице даны критические точки $t(\varepsilon, n.m)$ распределения $F_{n,m}$. Имеет место соотношение

$$t(1-\varepsilon, n, m) \cdot t(\varepsilon, m, n) \ = \ 1$$

| $\varepsilon \ = \ 0.005$ | | | | | | |
|---|---|---|---|---|---|---|
| $m \downarrow n \rightarrow$ | 4 | 5 | 10 | 20 | 24 | 60 |
| 4 | 23.15 | 22.46 | 20.97 | 20.17 | 20.03 | 19.61 |
| 5 | 15.56 | 14.94 | 13.62 | 12.90 | 12.78 | 12.40 |
| 10 | 7.34 | 6.87 | 5.85 | 5.27 | 5.17 | 4.86 |
| 20 | 5.17 | 4.76 | 3.85 | 3.32 | 3.22 | 2.92 |
| 24 | 4.89 | 4.49 | 3.59 | 3.06 | 2.97 | 2.67 |
| 60 | 4.14 | 3.76 | 2.90 | 2.39 | 2.29 | 1.96 |

| $\varepsilon \ = \ 0.01$ | | | | | | |
|---|---|---|---|---|---|---|
| $m \downarrow n \rightarrow$ | 4 | 5 | 10 | 20 | 24 | 60 |
| 4 | 15.98 | 15.52 | 14.55 | 14.02 | 13.93 | 13.65 |
| 5 | 11.39 | 10.97 | 10.05 | 9.55 | 9.47 | 9.20 |
| 10 | 5.99 | 5.64 | 4.85 | 4.40 | 4.33 | 4.08 |
| 20 | 4.43 | 4.10 | 3.37 | 2.94 | 2.86 | 2.61 |
| 24 | 4.22 | 3.89 | 3.17 | 2.74 | 2.66 | 2.40 |
| 60 | 3.65 | 3.34 | 2.63 | 2.20 | 2.11 | 1.84 |

| $\varepsilon \ = \ 0.05$ | | | | | | |
|---|---|---|---|---|---|---|
| $m \downarrow n \rightarrow$ | 4 | 5 | 10 | 20 | 24 | 60 |
| 4 | 6.39 | 6.26 | 5.96 | 5.80 | 5.77 | 5.67 |
| 5 | 5.19 | 5.05 | 4.73 | 4.56 | 4.53 | 4.43 |
| 10 | 3.48 | 3.33 | 2.98 | 2.77 | 2.74 | 2.62 |
| 20 | 2.87 | 2.71 | 2.35 | 2.12 | 2.08 | 1.95 |
| 24 | 2.78 | 2.62 | 2.25 | 2.03 | 1.98 | 1.84 |
| 60 | 2.52 | 2.37 | 1.99 | 1.75 | 1.70 | 1.53 |

одновременно. По условию,

$$\mathbf{P}(\eta < x) \ =$$
$$= \ \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} \int\limits_{*} \exp\{-\tfrac{1}{2} < A(\vec{z} - \vec{a}), (\vec{z} - \vec{a}) >\} \ d\vec{z},$$

где интеграл вычисляется по множеству $\{z : A^{\frac{1}{2}}(\vec{z} - \vec{a}) < x\}$. Сделаем замену $\vec{y} = A^{\frac{1}{2}}(\vec{z} - \vec{a})$. Тогда

$$< A(\vec{z} - \vec{a}), \ (\vec{z} - \vec{a}) > = < \vec{y}, \ \vec{y} >,$$

якобиан этой замены равен $\mid A \mid^{-\frac{1}{2}}$. Таким образом,

$$\mathbf{P}(\vec{\eta} < \vec{x}) \ = \int\limits_{\{\vec{y} < \vec{x}\}} \exp\{-\frac{1}{2} < \vec{y}, \ \vec{y} >\} \ d\vec{y},$$

что и утверждалось.

**Лемма 4.** $\vec{a} \ = \ \mathbf{M}\vec{\xi}.$

**Доказательство**. Очевидно, что если $\vec{\eta} \ = \ A^{\frac{1}{2}}(\vec{\xi} - \vec{a})$, то $\mathbf{M}\vec{\eta} \ = \ \vec{0}$ из соображений симметрии. К тому же, можно записать

$$\mathbf{M}\vec{\xi} \ = \ \mathbf{M}(A^{-\frac{1}{2}}\vec{\eta} \ + \ \vec{a}) \ = \ A^{-\frac{1}{2}}\mathbf{M}\vec{\eta} \ + \ \vec{a} \ = \ \vec{a}.$$

**Теорема 4.** *Характеристическая функция распределения с плотностью (8) имеет вид*

$$\varphi(\vec{t}) \ = \ \exp\{i < \vec{t}, \vec{a} > \ - \ \frac{1}{2} < M\vec{t}, \vec{t} >\}, \qquad \vec{t} \in \mathbf{R}^n,$$

*где $M \ = \ A^{-1}$, причем*

$$M_{i,j} \ = \ \mathbf{M}(\xi_i - a_i)(\xi_j - a_j), \quad i, j \ = \ 1, ..., n,$$

*т.е. $M$ - ковариационная матрица $\vec{\xi}$.*

С учетом этой теоремы далее распределение с плотностью (8) условимся обозначать $N(\vec{a}, M)$.

Докажем теорему. Пусть сначала $\vec{a} = \vec{0}$, $A = I$. Тогда

$$\varphi(\vec{t}\,) \;=\; \tfrac{1}{(2\pi)^{\frac{n}{2}}} \int\limits_{\mathbf{R}^n} \exp\{-\tfrac{1}{2} < \vec{x}, \vec{x} > \; + i < \vec{x}, \vec{x} >\} \, dx \quad =$$

$$= \prod_{j=1}^{n} \left( \tfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\{-\tfrac{1}{2} x_j^2 \; + \; i t_j x_j\} \, dx_j \right).$$

В скобках здесь записана характеристическая функция стандартного нормального распределения в точке $t_j$. Учитывая, что вид этой функции нам известен из теоремы 2, получим

$$\varphi(\vec{t}\,) \;=\; \prod_{j=1}^{n} \exp\{-\frac{1}{2} t_j^2\} \;=\; \exp\{-\frac{1}{2} < \vec{t}, \; \vec{t} >\}.$$

В общем случае требуемая формула выводится из леммы 3 и свойств многомерных характеристических функций.

С другой стороны, $\vec{\eta} \;=\; A^{\frac{1}{2}}(\vec{\xi} - \vec{a})$, откуда

$$\mathbf{cov}\vec{\xi} = \mathbf{M}(\vec{\xi} - \vec{a})(\vec{\xi} - \vec{a})^t \;=\; \mathbf{M}(A^{-\frac{1}{2}}\vec{\eta})(A^{-\frac{1}{2}}\vec{\eta})^t \;=\; A^{-\frac{1}{2}} \mathbf{cov}\vec{\eta} A^{-\frac{1}{2}}.$$

Но $\vec{\eta}$ имеет стандартное нормальное распределение, значит $\mathbf{cov}\vec{\eta} \;=\; I$, и, тем самым,

$$\mathbf{cov}\vec{\xi} \quad = \quad A^{-1},$$

что и утверждалось.

**Следствие 8.** *Если* $\vec{\xi} \;=\; (\xi_1, ..., \xi_n)$ *имеет распределение* $N(\vec{a}, M)$ *, то при произвольном* $j$

$$\xi_j \models N(a_j, \; M_{j,j}).$$

**Доказательство**. Заметим, что если

$$\vec{t}_{(j)} \;\equiv\; (0, ..., 0, t, 0, ..., 0) \;=\; t\vec{e_j},$$

то

$$\varphi_{\xi_j}(t) \;=\; \mathbf{M} e^{it\xi_j} \;=\; \mathbf{M} e^{i<\vec{t}_{(j)},\; \vec{\xi}>} \;=\; \varphi_{\vec{\xi}}(\vec{t}_{(j)}\,).$$

Привлекая теорему 4, имеем

$$\varphi_{\vec{\xi}}(\vec{t}_{(j)}\,) \;=\; \exp\{i < \vec{t}_{(j)},\; \vec{a} > \;-\; \tfrac{1}{2} < M\vec{t}_{(j)},\; \vec{t}_{(j)} >\} \;=$$
$$=\; \exp\{ita_j \;-\; \tfrac{1}{2} M_{j,j} t^2\}.$$

Следствие доказано. Более того, аналогичным образом нетрудно получить, что любой подвектор нормального вектора имеет нормальное распределение и выписать его параметры: вектор средних будет составлен из тех координат вектора средних $\vec{\xi}$, которые входят в исследуемый подвектор, а из $M$ вычеркиваются строки и столбцы с номерами тех координат, которые в наш подвектор не входят.

**Следствие 9.** *Если* $\vec{\xi} \;\models\; N(\vec{a}, M)$, *то*

$$\sum_{j=1}^{n} \xi_j \;\models\; N(\sum_j a_j,\; \sum_{i,j} M_{i,j}).$$

**Доказательство**. Пусть $S \;=\; \sum\limits_{j=1}^{n} \xi_j$. Заметим, что

$$\varphi_S(t) \;=\; \mathbf{M} e^{itS} \;=\; \mathbf{M} \exp(i < \vec{t},\; \vec{\xi} >) \;=\; \varphi_{\vec{\xi}}(\vec{t}),$$

где $\vec{t} = (t, t, ..., t)$, а значит, по теореме 4,

$$\varphi_S(t) \;=\; \exp(i < \vec{a},\; \vec{t} > \;-\; \tfrac{1}{2} < M\vec{t},\; \vec{t} >) \;=$$
$$=\; \exp\left\{it \sum_{j=1}^{n} a_j \;-\; \tfrac{t^2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} M_{i,j}\right\}.$$

Следствие доказано.

Это следствие часто формулируют так: сумма любых нормальных случайных величин нормальна. Такая формулировка ощибочна, т.к. в доказательстве существенно используется нормальность совместного распределения суммируемых случайных величин.

**Следствие 10.** *Если* $\vec{\xi} \models N(\vec{a}, M)$ *и* $\rho(\xi_i,\ \xi_j) = 0$ *при всех* $i \neq j$, *то координаты* $\vec{\xi}$ *независимы.*

**Доказательство**. Поскольку вне главной диагонали матрицы $M$ стоят $cov(\xi_i,\ \xi_j),\ i \neq j$, числители соответствующих коэффициентов корреляции, то ковариационная матрица диагональна. Значит,

$$\varphi_{\vec{\xi}}(\vec{t}) = \exp\left\{i \sum_{j=1}^{n} t_j a_j - \frac{t^2}{2} \sum_{j=1}^{n} M_{j,j}\right\} =$$
$$= \prod_{j=1}^{n} \varphi_j(t_j),$$

где $\varphi_j$ – характеристическая функция $N(a_j,\ M_{j,j})$. Но по следствию 8 $\xi_j \models N(a_j, M_{j,j}),\ j = 1, ..., n,$, откуда

$$\varphi_{\vec{\xi}}(\vec{t}) = \prod_{j=1}^{n} \varphi_{\xi_j}(t_j),$$

что и дает независимость координат.

К формулировке этого следствия можно сделать то же замечание, что и к следствию 9.

**Следствие 11.** *Если* $\vec{\xi} \models N(\vec{a}, M)$, *то*

$$< M^{-1}(\vec{\xi} - \vec{a}),\ (\vec{\xi} - \vec{a}) > \models \chi_n^2.$$

**Доказательство**. Если мы положим

$$\vec{\eta} = M^{-\frac{1}{2}}(\vec{\xi} - \vec{a}) = A^{\frac{1}{2}}(\vec{\xi} - \vec{a}) ,$$

то по лемме 3, $\vec{\eta} \models N(\vec{0}, I)$. Из следствий 8, 9 получается, что координаты $\vec{\eta}$ – независимые стандартные нормальные случайные величины, и значит,

$$< M^{-1}(\vec{\xi} - \vec{a}),\ (\vec{\xi} - \vec{a}) > = < \vec{\eta},\ \vec{\eta} > = \ \sum_{j=1}^{n} \eta_j^2 \ \models \ \chi_n^2.$$

# 7 Распределения некоторых полезных статистик

Пусть $X$ всюду в этом разделе обозначает выборку объема $n$ из распределения $N(a, \sigma^2)$. Напомним, что в этом случае можно смотреть на $X$ как на $n$-мерный случайный вектор с распределением $N(\vec{a},\ \sigma^2 I)$. При этом $\vec{a} \ = \ (a, ..., a)$ – вектор математических ожиданий, а $\sigma^2 I$ – ковариационная матрица $X$. Как всегда,

$$\bar{X} \ = \ \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad S^2 \ = \ \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{X})^2.$$

**Теорема 5.** $\frac{\bar{X}-a}{\sigma}\sqrt{n}$ *имеет стандартное нормальное распределение.*

Действительно, согласно следствию 9, $T \ \equiv \ n\bar{X}$ имеет распределение $N(na, n\sigma^2)$. По теореме 1 отсюда выводим, что

$$\frac{\bar{X} - a}{\sigma}\sqrt{n} \ = \ \frac{1}{\sigma\sqrt{n}}T \ - \ \frac{a\sqrt{n}}{\sigma} \ \models \ N(0, 1).$$

**Теорема 6.** *Статистика* $\frac{nS_a^2}{\sigma^2}$ , *где* $S_a^2 \ = \ \frac{1}{n}\sum_{j=1}^{n}(x_j - a)^2$ *имеет распределение* $\chi_n^2$.

Действительно (см. следствие 1), $\frac{x_j - a}{\sigma}$ имеет стандартное нормальное распределение. Поскольку $x_1, ..., x_n$ независимы, то по определению $\chi_n^2$ ,

$$\frac{nS_a^2}{\sigma^2} \;=\; \sum_{j=1}^{n}\left(\frac{x_j - a}{\sigma}\right)^2 \;\models\; \chi_n^2.$$

**Теорема 7.** *Статистики* $\bar{X}$ *и* $S^2$ *независимы, причем* $\frac{nS^2}{\sigma^2}$ *имеет распределение* $\chi_{n-1}^2$.

Докажем предварительно две леммы:

**Лемма 5.** *Пусть* $X \models N(\vec{0}, I)$ , $C$ *– ортогональная матрица,* $Y = CX$. *Тогда* $Y$ *имеет стандартное нормальное распределение, т.е. его координаты – стандартные нормальные независимые случайные величины.*

Для доказательства найдем характеристическую функцию $Y$:

$$\varphi_Y(\vec{t}) \;=\; \varphi_X(C^t\vec{t}) \;=\; \exp\{-\frac{1}{2} < C^t\vec{t}, C^t\vec{t} >\}.$$

Окончательно, учитывая, что $CC^t \;=\; I$, получаем

$$\varphi_Y(\vec{t}) \;=\; \exp\{-\frac{1}{2} < \vec{t}, \vec{t} >\} \;=\; \varphi_X(\vec{t}),$$

т.е. $Y$ и $X$ имеют одинаковые распределения.

**Лемма 6.** *Пусть* $X, Y$ *определены в лемме 5. Положим при* $r < n$

$$S(X) \;=\; \sum_{j=1}^{n} x_j^2 \;-\; y_1^2 \;-\; ... \;-\; y_r^2.$$

*Тогда* $S(X) \;\models \chi_{n-r}^2$ *и не зависит от* $y_1,\; ...,\; y_r$.

Действительно, заметим, что

$$\sum_{j=1}^{n} x_j^2 \; = < X, X > = < CX, CX > = \; \sum_{j=1}^{n} y_j^2.$$

Поэтому $S(X) \; = \; \sum\limits_{j=r+1}^{n} y_j^2$. Поскольку, согласно лемме 5,

$$y_1, ..., y_n \; \models \; N(0,1)$$

и независимы, то доказательство леммы завершено.

Докажем теперь теорему. Обозначим

$$z_j \; = \; \frac{x_j - a}{\sigma} \; \models \; N(0,1), \quad j = 1, .., n.$$

Заметим, что

$$\begin{gathered}
\tfrac{\bar{X}-a}{\sigma} \; = \; \tfrac{1}{n} \sum_{j=1}^{n} z_j \; = \; \bar{Z}; \\
\tfrac{S^2}{\sigma^2} \; = \; \tfrac{1}{n} \sum_{j=1}^{n} (z_j - \bar{Z})^2 \; = \; \tfrac{1}{n} \sum_{j=1}^{n} z_j^2 \; - \; (\bar{Z})^2; \\
\tfrac{nS^2}{\sigma^2} \; = \; \sum_{j=1}^{n} z_j^2 \; - \; (\sqrt{n}\bar{Z})^2.
\end{gathered}$$

Положим $\vec{C}_1 \; = \; (\frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}})$. Так как $\|\vec{C}_1\| \; = \; 1$, то можно найти векторы $\vec{C}_2, ..., \vec{C}_n$, дополняющие $\vec{C}_1$ до ортонормированного базиса в $\mathbf{R}^n$. Матрица $C$, по строкам которой размещены координаты найденных векторов, ортогональна. Если $Y \; = \; CZ$, то выполнены условия леммы 6, причем $y_1 \; = \; \sqrt{n}\bar{Z}$. Тем самым, $\frac{nS^2}{\sigma^2} \models \chi^2_{n-1}$ и не зависит от $\bar{X} \; = \; \frac{\sigma y_1}{\sqrt{n}} + a$. Теорема доказана.

**Следствие 12.**

$$\frac{\bar{X} - a}{S} \sqrt{n-1} \; \models \; T_{n-1}.$$

Это утверждение немедленно следует из теорем 5, 6 и представления

$$\frac{\bar{X}-a}{S}\sqrt{n-1} \;=\; \frac{\frac{\bar{X}-a}{\sigma}\sqrt{n}}{\sqrt{\frac{1}{n-1}\;\frac{nS^2}{\sigma^2}}}.$$

# 8 Точные доверительные интервалы для параметров нормального распределения

Говорят, что статистики $\theta^{\pm}(X)$ образуют *границы точного доверительного интервала* для $\theta$ уровня $1-\varepsilon$, если

$$\mathbf{P}(\theta^{-}(X) \;\le\; \theta \;\le\; \theta^{+}(X)) \;=\; 1-\varepsilon.$$

Утверждения предыдущего раздела позволяют нам построить такие интервалы в случае, когда $X$ – выборка из $N(a,\sigma^2)$.

## 8.1 Доверительные интервалы для математического ожидания

Согласно следствию 12 случайная величина $\eta \;=\; \frac{\bar{X}-a}{S}\sqrt{n-1}$ имеет распределение Стьюдента с $n-1$ степенью свободы. Найдем по заданному $\varepsilon$ из таблиц $T_{n-1}$ на странице 16 такое $\tau_\varepsilon$, что

$$\mathbf{P}(\mid \eta \mid\ge \tau_\varepsilon) \;=\; \varepsilon.$$

Это означает, что неравенство $\mid \eta \mid< \tau_\varepsilon$ будет выполнено с вероятностью $1-\varepsilon$. Разрешая последнее неравенство относительно $a$, получим

$$\mathbf{P}\left(\bar{X} \;-\; \frac{S\tau_\varepsilon}{\sqrt{n-1}} \;\le\; a \;\le \bar{X} \;+\; \frac{S\tau_\varepsilon}{\sqrt{n-1}}\right) \;=\; 1-\varepsilon,$$

а следовательно, границы искомого доверительного интервала задаются формулой

$$a^{\pm} \;=\; \bar{X} \;\pm\; \frac{S\tau_\varepsilon}{\sqrt{n-1}}. \tag{9}$$

При расчете этих границ мы не пользуемся никакой информацией кроме той, что содержится в самой выборке. Естественно, что если мы знаем, например, чему равна дисперсия, то ширину интервала (9) можно уменьшить, используя заданное значение дисперсии вместо оценки $S$. Это можно сделать, опираясь на теорему 5.

По таблицам нормального распределения на странице 9 найдем такое $t_\varepsilon$, что

$$\Phi(t_\varepsilon) \;=\; \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty} t_\varepsilon \exp\left\{-\,\frac{y^2}{2}\right\}\, dy \;=\; 1-\frac{\varepsilon}{2},$$

Тогда по следствию 3, $\Phi(-t_\varepsilon) \;=\; \frac{\varepsilon}{2}$, и значит, поскольку нам известно, что $\frac{\bar{X}-a}{\sigma}\sqrt{n} \;\models\; N(0,1)$, то имеет место равенство

$$\mathbf{P}\left(\frac{\mid \bar{X}-a \mid}{\sigma}\sqrt{n} \;<\; t_\varepsilon\right) \;=\; \Phi(t_\varepsilon) \;-\; \Phi(-t_\varepsilon) \;=\; 1-\varepsilon.$$

Решая неравенство под знаком вероятности относительно $a$, получим

$$\mathbf{P}\left(\bar{X} \;-\frac{\sigma t_\varepsilon}{\sqrt{n}} \;<\; a \;< \bar{X} \;+\frac{\sigma t_\varepsilon}{\sqrt{n}}\right) \;=\; 1-\varepsilon,$$

или

$$a^{\pm} \;=\; \bar{X} \;\pm \frac{\sigma t_\varepsilon}{\sqrt{n}}. \tag{10}$$

Интервал (10) явно уже, чем (9). Итак, если $\sigma$ откуда-то нам известно, то границы доверительного интервала находим по формуле (10), иначе – по формуле (9).

## 8.2 Доверительные интервалы для дисперсии

Если мы введем обозначение $\chi \equiv \frac{nS^2}{\sigma^2}$, то согласно теореме 7, $\chi \models \chi^2_{n-1}$. По таблицам распределения $\chi^2$ на стр 14 найдем такие $\tau_\varepsilon^\pm$, что

$$\mathbf{P}(\tau_\varepsilon^- < \chi < \tau_\varepsilon^+) = 1 - \varepsilon.$$

Это можно сделать многими способами, но обычно исходят из условий

$$\mathbf{P}(\chi > \tau_\varepsilon^-) = 1 - \frac{\varepsilon}{2}, \qquad \mathbf{P}(\chi > \tau_\varepsilon^+) = \frac{\varepsilon}{2}.$$

Разрешая двойное неравенство под знаком вероятности относительно $\sigma^2$, имеем

$$\mathbf{P}\left(\frac{nS^2}{\tau_\varepsilon^+} < \sigma^2 < \frac{nS^2}{\tau_\varepsilon^-}\right) = 1 - \varepsilon,$$

или

$$(\sigma^2)^\pm = \frac{nS^2}{\tau_\varepsilon^\mp}. \tag{11}$$

Опять-таки, если известно, чему равно $a$, то вместо теоремы 7 мы можем воспользоваться теоремой 6, определить $t_\varepsilon^\pm$ на этот раз по таблице $\chi^2_n$ как критические точки уровней $1 - \frac{\varepsilon}{2}$ и $\frac{\varepsilon}{2}$ соответственно, и получить

$$\mathbf{P}\left(\frac{nS_a^2}{t_\varepsilon^+} < \sigma^2 < \frac{nS_a^2}{t_\varepsilon^-}\right) = 1 - \varepsilon.$$

Границами доверительного интервала здесь будут

$$(\sigma^2)^\pm = \frac{nS_a^2}{t_\varepsilon^\mp}. \tag{12}$$

Ошибка в определении $a$ для расчета (12) чревата гораздо более серьезными последствиями, чем неверно определенное $\sigma$ в (10). Там ошибка могла сказаться лишь на ширине интервала с неизменным центром $\bar{X}$ для (9) и (10) , а интервалы (11) и (12) могут даже оказаться непересекающимися.

## 9 Критерий $\chi^2$ : определение

Изначально этот критерий нацелен на проверку простой гипотезы о том, что наблюдаемая случайная величина $\xi$ имеет известную заранее функцию распределения $F$. Предположим, что группы $\Delta_1, ..., \Delta_r$ построены так, что $\xi$ может принимать лишь значения, попадающие в одну из этих групп. Чаще всего это интервалы числовой прямой. Обсуждение способов построения подобных интервалов отложим до раздела 11. Проверяемая гипотеза подменяется предположением о том, что значения из $j$-й группы случайная величина принимает с вероятностями $p_j,\ j = 1, ..., r$. В ситуации, когда функция распределения известна,эти вероятности нетрудно вычислить (подробности в также разделе 11). Конечно же, предполагается, что

$$\sum_{j=1}^{r} p_j \quad = \quad 1. \tag{13}$$

Пусть $\nu_j$ - число элементов выборки $X$ объема $n$, попавших в группу $\Delta_j$. Очевидно, что

$$\sum_{j=1}^{r} \nu_j \quad = \quad n \tag{14}$$

*Статистикой* $\chi^2$ называется величина

$$\chi^2 \ = \ \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j}.$$

Если проверяемая гипотеза справедлива, то $np_j$ представляет собой ожидаемое число элементов выборки, попавших в $\Delta_j$, а $\frac{(\nu_j - np_j)^2}{np_j}$ – относительное отклонение экспериментальных данных от ожидаемых теоретически. Таким образом, гипотеза должна быть принята, если $\chi^2$ не слишком велико.

Естественным является критерий, основанный на $\chi^2$, и с точки зрения теории. Если наша гипотеза имеет место, то, как нетрудно понять, функция правдоподобия имеет вид

$$L \;=\; L(X, p_1, ..., p_r) \;=\; \frac{n!}{\nu_1! ... \nu_r!}\, p_1^{\nu_1} ... p_k^{\nu_r}.$$

Если же $\tilde{p_1}, ..., \tilde{p_r}$ – альтернативные значения $p_j$ , то

$$\tilde{L} \;=\; L(X, \tilde{p_1}, ..., \tilde{p_r}) \;=\; \frac{n!}{\nu_1! ... \nu_r!}\, \tilde{p_1}^{\nu_1} ... \tilde{p_k}^{\nu_r}.$$

Оценим параметры альтернативы по методу максимума правдоподобия:

$$l \;=\; \ln L \;= \sum_{j=1}^{r} \nu_j \ln p_j \;+\; \ln \frac{n!}{\nu_1! ... \nu_r!}.$$

В силу формулы (13) имеем

$$l'_j \;=\; \frac{\nu_j}{p_j} \;-\; \frac{\nu_r}{p_r}, \quad j = 1, ..., r-1.$$

Приравнивая производные нулю, складывая все полученные уравнения и привлекая (14), получаем

$$n \;=\; \sum_{j=1}^{r} \nu_j \;=\; \frac{\nu_r}{p_r} \sum_{j=1}^{r} p_j \;=\; \frac{\nu_r}{p_r},$$

откуда

$$\hat{p_r} \;=\; \frac{\nu_r}{n}$$

и

$$\hat{p_j} \; = \; \hat{p_r} \cdot \frac{\nu_j}{\nu_r} \; = \; \frac{\nu_j}{n}, \quad j = 1, ..., r.$$

Окончательно,

$$\frac{\tilde{L}}{L} \; = \; \prod_{j=1}^{r} \left( \frac{\nu_j}{np_j} \right)^{\nu_j}.$$

Рассматривая логарифм этого отношения и вновь используя (13, 14), имеем

$$\ln \tfrac{\tilde{L}}{L} \; = \; \textstyle\sum_{j=1}^{r} \nu_j \ln\left(1 \; + \; \frac{\nu_j - np_j}{np_j}\right) \asymp \textstyle\sum_{j=1}^{r} \nu_j \frac{\nu_j - np_j}{np_j} \; =$$
$$= \; \textstyle\sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} \; + \; \textstyle\sum_{j=1}^{r} (\nu_j - np_j) \; = \; \chi^2.$$

Итак, с точностью до членов второго порядка малости, $\chi^2$ совпадает с критерием, основанным на отношении правдоподобия.

## 10 Критерий $\chi^2$: теоретическое обоснование

В этом пункте будет доказана

**Теорема 8.** *При* $n \to \infty$

$$\chi^2 \; = \; \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} \; \xrightarrow{d} \; \chi^2_{r-1}.$$

**Доказательство**. Рассмотрим векторы $\mathbf{p} \; = (p_1, ..., p_r)$, $\nu \; = \; (\nu_1, ..., \nu_r)$. Очевидно, что $\nu$ имеет полиномиальное распределение с параметрами $n$, $\mathbf{p}$, т.е., если $\mathbf{a}$ – вектор с целочисленными неотрицательными координатами $a_1, ..., a_r$, такой, что сумма этих координат равна $n$, то

$$\mathbf{P}(\nu = \mathbf{a}) \; = \; \frac{n!}{a_1! ... a_r!} \, p_1^{a_1} ... p_r^{a_r}.$$

**Лемма 7.** *Характеристическая функция* $\nu$ *имеет вид*

$$\varphi_\nu(t) \;=\; \left(\sum_{j=1}^{r} p_j e^{it_j}\right)^n, \quad t \in \mathbf{R}^r.$$

Действительно,

$$\varphi_\nu(t) \;=\; \mathbf{M}e^{i<t,\nu>} \;=\; \Sigma_{\mathbf{a}} e^{i<t,\mathbf{a}>}\mathbf{P}(\nu = \mathbf{a}) \;=$$
$$=\; \Sigma_{\mathbf{a}} \tfrac{n!}{a_1!\ldots a_r!} \left(p_1 e^{it_1}\right)^{a_1} \ldots \left(p_r e^{it_r}\right)^{a_r} \;=\; \left(\textstyle\sum\limits_{j=1}^{r} p_j e^{it_j}\right)^n.$$

Лемма доказана. Положим теперь

$$\nu^* \;=\; \frac{\nu - n\mathbf{p}}{\sqrt{n}} \;=\; \frac{1}{\sqrt{n}}\nu \;-\; \sqrt{n}\mathbf{p}.$$

Из леммы следует, что

$$\varphi_{\nu^*}(t) \;=\; \exp\{-i\sqrt{n} < \mathbf{p}, t >\} \left(\textstyle\sum_{j=1}^{r} p_j e^{it_j/\sqrt{n}}\right)^n,$$
$$l(t) = \ln \varphi_\nu(t) \;=$$
$$=\; -i\sqrt{n} < \mathbf{p}, t > + n \ln\left(1 + \textstyle\sum_{j=1}^{r} p_j (e^{it_j/\sqrt{n}} - 1)\right).$$

Имея ввиду, что

$$e^{it_j/\sqrt{n}} - 1 \;=\; \tfrac{it_j}{\sqrt{n}} \;-\; \tfrac{t_j^2}{2n} \;+\; o\left(\tfrac{1}{n\sqrt{n}}\right);$$
$$\ln(1+\epsilon) \;=\; \epsilon \;-\; \tfrac{\epsilon^2}{2} \;+\; o(\epsilon^2),$$

выводим

$$l(t) \;=\; -\tfrac{1}{2} \textstyle\sum\limits_{j=1}^{r} t_j^2 p_j \;+\; \tfrac{1}{2} \left(\textstyle\sum\limits_{j=1}^{r} t_j p_j\right)^2 \;+\; o\left(\tfrac{1}{\sqrt{n}}\right) \;=$$
$$=\; -\tfrac{1}{2} < \tilde{A}t,\, t > \;+ o\left(\tfrac{1}{\sqrt{n}}\right),$$

где матрица $\tilde{A}$ размерами $r \times r$ имеет элементы

$$\tilde{A}_{i,j} \;=\; \begin{cases} p_j(1-p_j), & i = j; \\ -\,p_i p_j, & i \neq j, \end{cases} \;.\; i, j = 1, \ldots, r.$$

Итак,

$$\varphi_{\nu^*}(t) \ \to \ \exp\{-\frac{1}{2} < \tilde{A}t,\ t >\}, \tag{15}$$

но, к сожалению, $\tilde{A}$ – вырожденная матрица. Это имеет место за счет зависимости координат $\nu^*$:

$$\sum_{j=1}^{r} \nu_j^* \ = \ 0. \tag{16}$$

В силу замечания, сделанного после следствия 8 из (15) выводится, что

$$\varphi_{\tilde{\nu}}(t) \ \to \ \exp\{< At,\ t >\},$$

где $\tilde{\nu} \ = \ (\nu_1^*, ..., \nu_{r-1}^*)$, а $A$ получается из $\tilde{A}$ вычеркиванием последней строки и последнего столбца. Отсюда

$$\tilde{\nu} \ \xrightarrow{d} \ N(0, A).$$

Матрица $A$ обратима, причем обратная к ней матрица $B$ состоит их элементов

$$B_{i,j} \ = \ \begin{cases} \frac{1}{p_i} \ + \frac{1}{p_r}, & i = j; \\ \frac{1}{p_r}, & i \neq j, \end{cases} \qquad i, j = 1, ..., r-1.$$

Проведем следующие выкладки

$$< B\tilde{\nu},\ \tilde{\nu} >= \ \Sigma_{i,j}\ B_{i.j}\nu_i^*\nu_j^* \ = \ \sum_{j=1}^{r-1} \left(\frac{1}{p_j} + \frac{1}{p_r}\right)(\nu_j^*)^2 \ +$$

$$+ \ \frac{1}{p_r}\Sigma_{i\neq j}\nu_i^*\nu_j^* \ = \ \tfrac{1}{p_r}\left(\sum_{j=1}^{r-1}\nu_j^*\right)^2 + \ \sum_{j=1}^{r-1}\frac{(\nu_j - np_j)^2}{np_j}.$$

Поскольку (см (16))

$$\frac{1}{p_r}\left(\sum_{j=1}^{r-1}\nu_j^*\right)^2 \ = \ \frac{(\nu_r^*)^2}{p_r} \ = \ \frac{(\nu_r - np_r)^2}{np_r},$$

то по следствию 11,

$$\chi^2 = < B\tilde{\nu},\, \tilde{\nu} > \xrightarrow{d} \chi^2_{r-1}.$$

Конечно же, гипотеза о принадлежности выборки определенному распределению встречается относительно реже, чем задача принадлежности некоторому параметрическому семейству распределений – нормальному, пуассоновскому и т.п. В этих случаях по выборке дополнительно приходится оценивать неизвестные параметры распределений. Соответствующий результат сформулируем следующим образом:

**Теорема 9.** *(Фишер) Если при расчете статистики $\chi^2$ $k$ неизвестных параметров распределения заменены их оценками по методу максимума правдоподобия, то в достаточно широких предположениях предельным распределением $\chi^2$ будет $\chi^2_{r-k-1}$.*

Доказательства этой теоремы мы здесь приводить не будем.

## 11 Применение $\chi^2$ к гипотезе нормальности

Рассмотрим сначала подготовку групп для работы критерия $\chi^2$. Если эти группы не определены естественным образом, и данные имеют числовой характер, то прежде всего определим так называемый *размах выборки* $T = \mid X_{(n)} - X_{(1)} \mid$. Затем выберем число групп $r$. Для первого приближения рекомендуется пользоваться *формулой Стерджеса*:

$$r = [\log_2 n] + 1.$$

Определим длину типичного интервала (шаг) $h = T/r$. Построим концы групп-интервалов по формулам

$$z_1 = -\infty,\ z_k = X_{(1)} + (k-1)h\ (2 \le k \le r),\ z_{r+1} = +\infty. \quad (17)$$

Положим $\Delta_j = [z_j,\ z_{j+1})$, $j = 1, ..., r$. Каждый из построенных интервалов должен ОБЯЗАТЕЛЬНО содержать **от 3 до 19** выборочных значений. Если это условие нарушается, то интервалы следует изменить, передвигая границы $z_k$, или изменяя количество групп.

Осталось теперь выяснить, как вычислять $p_j$. Пусть проверяется гипотеза о том, что выборка взята из распределения с функцией $F$. Тогда

$$p_j = F(z_{j+1}) - F(z_j), \quad j = 1, ..., r$$

(естественно, после того, как группы $\Delta_1, ..., \Delta_r$ уже окончательно определены).

Рассмотрим гипотезу нормальности: пусть мы хотим проверить гипотезу о том, что выборка взята из распределения $N(a, \sigma^2)$ с некоторыми неизвестными параметрами. По теореме Фишера можно взять

$$a = \hat{a} = \bar{X}, \quad \sigma^2 = \hat{\sigma}^2 = S^2,$$

что уменьшит число степеней свободы на два. Построим границы групп по формулам (17) и положим

$$p_j = \mathbf{P}(z_j \le \xi < z_{j+1}) = \mathbf{P}\left(\frac{z_j - \bar{X}}{S} \le \frac{\xi - a}{\sigma} < \frac{z_{j+1} - \bar{X}}{S}\right) =$$

$$= \Phi\left(\frac{z_{j+1} - \bar{X}}{S}\right) - \Phi\left(\frac{z_j - \bar{X}}{S}\right), \qquad j = 1, ..., r. \qquad (18)$$

Дальнейшие результаты вычислений оформляем таблицей:

**Проверка гипотезы нормальности**

| N | характеристика | как определяется |
|---|---|---|
| 1 | $z_j$ | задаются заранее (17) |
| 2 | $(z_j - \bar{X})/S$ | по строке 1 |
| 3 | $\Phi((z_j - \bar{X})/S)$ | по таблице 3 $\Phi$ |
| 4 | $p_j$ | по формулам (18) и строке 1 |
| 5 | $np_j$ | по строке 4 |
| 6 | $\nu_j$ | по выборке и группам |
| 7 | $(\nu_j - np_j)^2$ | по строкам 5 и 6 |
| 8 | $(\nu_j - np_j)^2/np_j$ | по строкам 5 и 7 |

Величина статистики $\chi^2$ определяется как сумма элементов 8-й строки. После расчета сравниваем ее с критической точкой, определенной по таблице 4 для $\chi^2_{r-3}$ по заданному $\varepsilon$. Если критический уровень не превзойден, то отвергать гипотезу нормальности нет оснований.

## 12 Задачи с несколькими выборками

В этом разделе рассмотрены применения критерия $\chi^2$ к проверке гипотез независимости и однородности выборок.

Рассмотрим две выборки $X$ и $Y$ одинакового объема $n$. Гипотеза независимости состоит в том, что эти выборки взяты из независимых распределений. Тем самым высказываемая гипотеза относится к совместному распределению $X$ и $Y$, и фактически надо считать, что мы имеем выборку объема $n$ из двумерного распределения, в которой элементы выборки $X$ образуют первые координаты, а элементы $Y$ – вторые.

Ясно,что если мы будем группировать эти двумерные данные, то группы будут представлять собой прямоугольники.

Пусть мы имеем $r \times s$ прямоугольников-групп, и каждой из групп, обозначаемых $\Delta_{i,j}$, соответствует теоретическая вероятность

$$p_{i,j} \ = \ \mathbf{P}((x, y) \in \Delta_{i,j}), \ i = 1, ..., r, \ j = 1, ..., s.$$

В этом случае

$$\chi^2 \ = \ \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\nu_{i,j} - np_{i,j})^2}{np_{i,j}} \ \xrightarrow{d} \ \chi^2_{rs-1},$$

где $\nu_{i,j}$ – число элементов двумерной выборки, попавших в $\Delta_{i,j}$.

Высказываемая гипотеза независимости заменяется теперь на утверждение о том, что

$$p_{i,j} \ = \ p_i q_j, \ \ p_i \ = \ \mathbf{P}(x \in \Delta_i), \ \ q_j \ = \ \mathbf{P}(y \in \Lambda_j),$$

если $\Delta_{i,j} \ = \ \Delta_i \times \Lambda_j, \ i = 1, ..., r, \ j = 1, ..., s.$

Обычно $\Delta_{i,j}$ строят так: образуем для выборки $X$ интервалы $\Delta_1, ..., \Delta_r$, а для выборки $Y$ интервалы $\Lambda_1, ..., \Lambda_s$ как это было сделано в предыдущем разделе и положим $\Delta_{i,j} \ = \ \Delta_i \times \Lambda_j$. Параметры $p_i, q_j$, являющиеся неизвестными, заменим их оценками максимума правдоподобия:

$$\hat{p}_i \ = \ \frac{\nu_i}{n}, \ \hat{q}_j \ = \ \frac{\mu_j}{n}, \ i = 1, ..., r, \ j = 1, ..., s,$$

где $\nu_i$ – число элементов $X$ в $\Delta_i$, $\mu_j$ – число элементов $Y$ в $\Lambda_j$. После несложных преобразований получим, что если гипотеза справедлива, то

$$\chi^2 \ = \ \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\nu_{i,j} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j} \ = \ n\left(\sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\nu_{i,j}^2}{\nu_i \mu_j} - 1\right), \qquad (19)$$

и, согласно теореме Фишера,

$$\chi^2 \ \xrightarrow{d} \ \chi^2_{rs-(r-1)-(s-1)} \ = \ \chi^2_{(r-1)(s-1)}.$$

Обычно данные для проверки гипотезы независимости помещают в таблицу, называемую *таблицей сопряженности*:

**Таблица сопряженности**

| $Y \setminus X$ | $\Delta_1$ | ... | $\Delta_r$ | суммы строк |
|---|---|---|---|---|
| $\Lambda_1$ | | | | $\mu_1$ |
| $\vdots$ | | $\nu_{i,j}$ | | $\vdots$ |
| $\Lambda_s$ | | | | $\mu_s$ |
| суммы столбцов | $\nu_1$ | ... | $\nu_r$ | |

Числа $\nu_{i,j}$ находят по выборке, проверяя каждую из двумерных выборочных точек: если $x$ попал в $i$-й интервал, а $Y$ в $j$-й, то $\nu_{i,j}$ увеличивают на единицу (таким образом, во всей таблице сопряженности появятся $n$ единиц). После вычислений сумм строк и столбцов, рассчитывают $\chi^2$ по формуле (19) и, сравнивая с критическим значением по таблице $\chi^2_{(r-1)(s-1)}$ делают вывод о согласованности гипотезы опытным данным.

Далее рассмотрим гипотезу однородности выборок. Эта гипотеза состоит в том, что выборки $X,\ Y$, имеющие на этот раз возможно различные объемы $n,\ m$ соответственно, берутся из одного распределения.

Объединим две наши выборки в одну и будем считать, что мы наблюдаем двумерный вектор $(Z,\ v)$, где $Z$ пробегает все значения объединенной выборки объема $n+m$ и $v=1$, если соответствующий $Z$ взят из выборки $X$. Иначе $v=2$. Если выборки однородны, то $Z$ не зависит от $v$. Итак, мы пришли к гипотезе независимости.

Построим теперь таблицу сопряженности. С этой целью по *объединенной* выборке построим группы $\Delta_1, ..., \Delta_r$ (см. предыдущий раздел). Пусть $\psi_i,\ \varphi_i$ соответственно – количества элементов первой и второй выборок, попавших в $i$-ю группу. Таблица сопряженности имеет вид

**Проверка гипотезы однородности**

| $\upsilon \setminus Z$ | $\Delta_1$ | $\Delta_2$ | ... | $\Delta_r$ | суммы строк |
|---|---|---|---|---|---|
| $\upsilon = 1$ | $\psi_1$ | $\psi_2$ | ... | $\psi_r$ | $n$ |
| $\upsilon = 2$ | $\varphi_1$ | $\varphi_2$ | ... | $\varphi_r$ | $m$ |
| суммы столбцов | $\nu_1$ | $\nu_2$ | ... | $\nu_r$ | $n+m$ |

Здесь

$$\sum_{i=1}^{r} \psi_i = n, \quad \sum_{j=1}^{r} \varphi_j = m, \quad \psi_i + \varphi_i = \nu_i.$$

Согласно формуле (19),

$$\chi^2 = (n+m)\left(\sum_{i=1}^{r} \frac{\psi_i^2}{n\nu_i} + \sum_{j=1}^{r} \frac{\varphi_i^2}{m\nu_j} - 1\right) =$$
$$= (n+m)\left(\sum_{i=1}^{r} \frac{1}{\nu_i}\left(\frac{\psi_i^2}{n} + \frac{\varphi_i^2}{m}\right) - 1\right), \qquad (20)$$

и имеем $\chi^2_{(r-1)(2-1)} = \chi^2_{r-1}$ в качестве предельного распределения. После вычисления значения $\chi^2$ по формуле (20) остается, как обычно, сравнить его с критической точкой $\chi^2_{r-1}$.

## 13 Гипотезы равенства дисперсий и средних

Пусть $X$ – выборка объема $n$ из $N(a_1, \sigma_1^2)$ , $Y$ – объема $m$ из $N(a_2, \sigma_2^2)$. Установим сначала критерий для проверки гипотезы $\sigma_1 = \sigma_2$. Обозначим общее (неизвестное) значение дисперсий через $\sigma^2$.

Из теоремы 6 и определения распределения Фишера следует, что если гипотеза справедлива, то

$$S = \frac{S_X^2}{S_Y^2} = \frac{m(nS_X^2/\sigma^2)}{n(mS_Y^2/\sigma^2)}$$

имеет распределение $F_{n-1,m-1}$. Дальнейшие рассуждения стандартны. По таблице 5 $F$-распределения находим критические точки $f^-$ и $f^+$ уровней $1-\frac{\varepsilon}{2}$, $\frac{\varepsilon}{2}$ соответственно. Если рассчитанная по выборке величина $S$ попала в интервал $(f^-,\ f^+)$, то гипотеза верна на уровне $1-\varepsilon$.

Проверку гипотезы $a_1 \ = \ a_2$ обычно связывают с предварительной проверкой равенства дисперсий (см.[2]). Дело в том, что не существует симметричного относительно элементов выборок $X$, $Y$ критерия проверки этой гипотезы, не использующего совпадения их дисперсий. По этому поводу см. [4, с. 192–194]. Но там же приведен следующий просто устроенный несимметричный критерий, который можно применять без каких-либо предварительных проверок.

Пусть $n \ \leq m$. Построим

$$u_i \ = \ x_i \ - \ \sqrt{\frac{n}{m}}\, y_i, \qquad i=1,...,n$$

и определим

$$\bar{U} \ = \frac{1}{n}\sum_{i=1}^{n} u_i; \quad S_U^2 \ = \ \frac{1}{n}\sum_{i=1}^{n}(u_i-\bar{U})^2.$$

Тогда, если только наша гипотеза имеет место, то

$$T \ = \ \frac{\bar{X}-\bar{Y}}{S_U}\sqrt{n-1} \ \models T_{n-1}. \tag{21}$$

Действительно, согласно теореме 1,

$$u_i \ \models \ N\left(a(1-\sqrt{\frac{n}{m}}),\ \sigma_1^2+\frac{n}{m}\sigma_2^2\right), \quad i=1,...,n.$$

Отсюда по теореме 6,

$$\frac{S_U^2}{\sigma_1^2/n \ + \ \sigma_2^2/m} \ \models \ \chi^2_{n-1}.$$

С другой стороны, очевидно, что

$$\bar{X} \ - \bar{Y} \ \models \ N(0, \ \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}),$$

поэтому в силу определения распределения Стьюдента,

$$T \ \stackrel{d}{=} \ \frac{N(0,1)}{\sqrt{\frac{1}{n-1}\chi^2_{n-1}}} \ = \ T_{n-1},$$

что и утверждалось.

Итак, рассчитав $T$ по формуле (21) и сравнив его с двусторонней критической точкой для $T_{n-1}$ по таблице 5, можно сделать вывод о соответствии гипотезы равенства средних опытным данным.

# 14 О статистических расчетах с применением компьютера

В этом небольшом разделе приведен необходимый минимум справочного материала, необходимого при проведении расчетов, описанных в предыдущих разделов в пакете Microsoft Excel. Конечно же, имеется большое число прикладных компьютерных пакетов, изначально ориентированных на статистические расчеты: Statistica, SPSS, Stata, Origin и многие другие. Они очень удобны, но предоставляют по исходным данным сразу готовый результат. Нам же надо понять, почему этот результат именно таков, и как все это работает. По перечисленным причинам использование упомянутых пакетов при выполнении расчетного задания не приветствуется.

Сначала приведем перечень самых простых статистических функций без излишних подробностей. Доступ к ним можно получить через меню "Вставка функции", "Статистические".

- СРЗНАЧ – вычисляет $\overline{X}$ в наших обозначениях;
- ДИСПР, ДИСП – соответственно, $S^2$, $\frac{n}{n-1}S^2$. Имеющиеся там же функции ДИСПА, ДИСПРА работают аналогично первым двум, если в массиве ваших данных нет логических переменных и текста;
- КОРРЕЛ – вычисляет выборочный коэффициент корреляции;

Следующие несколько функций связаны с вычислением квантилей и критических точек распределений. Они доступны через те же разделы меню Excel, что и предыдущие.

- НОРМСТРАСП – вычисляет значение $\Phi(x)$;
- НОРМСТОБР – по заданному $\varepsilon$ находит квантиль стандартного нормального распределения уровня $\varepsilon$;
- СТЬЮРАСПОБР – по заданному $\varepsilon$ и числу степеней свободы находит двустороннюю критическую точку распределения Стьюдента;
- ХИ2ОБР – находит одностороннюю критическую точку заданного уровня (не квантиль!) хи-квадрат распределения по его числу степеней свободы;
- FРАСПОБР – позволяет вычислить одностороннюю критическую точку распределения Фишера по задаваемому уровню ("вероятность") и паре чисел, задающих количества степеней свободы числителя и знаменателя;
- ГАММАОБР – с помощью этой функции можно находить критические точки гамма-распределений.

Конечно же, в разделе "Статистические функции"имеется и множество других полезных функций. Для вычислений же векторных характеристик, например, ковариационных и корреляционных матриц, дополнительные возможности заложены в так называемом "пакете анализа данных". Он доступен через меню "Сервис", "Анализ данных". Этот пакет не устанавливается автоматически при установке Excel, но для пользования им дистрибутив Microsoft Office не требуется. Достаточно в меню "Сервис", "Надстройки"поставить галочку напротив пункта "Пакет анализа", после чего он будет готов к работе.

Литературный список

1. Агапов Г.И. Задачник по теории вероятностей. - М.:Высшая школа, 1986.- 80 с.

2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике.- М.:Высшая школа, 1979.- 400 с.

3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.:Наука, 1983.- 416 с.

4. Кендалл М., Стьюарт А. Статистические выводы и связи.- М.:Наука, 1973.- 900 с.