Стандартные распределения и их квантили

Стандартные распределения

В статистике, эконометрике и других сферах человеческих знаний очень часто используются стандартные распределения.

В частности, они используются для проверки гипотез и построения доверительных интервалов.

Самыми распространенными являются:

1. Нормальное распределение z .

2. Распределение **Пирсона** или **хи-квадрат** ($\chi_n{}^2$).

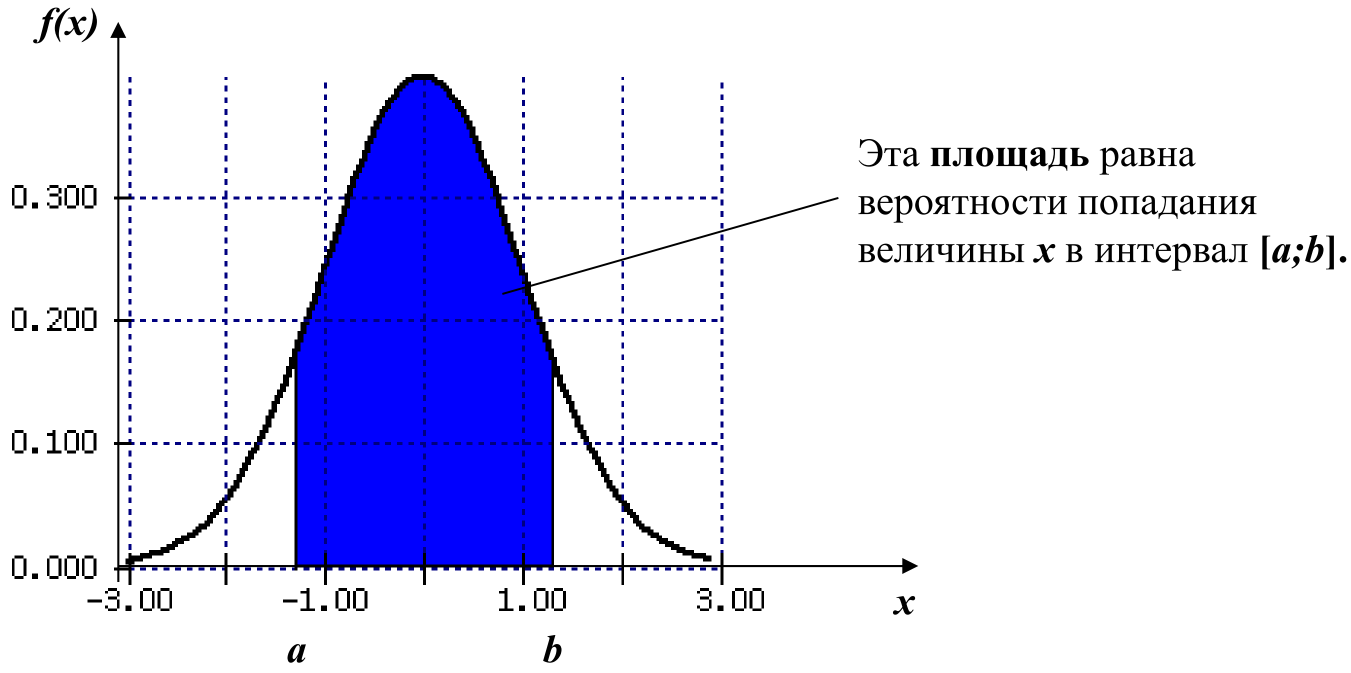
3. Распределение **Стьюдента (**t_n**)**.

4. Распределение **Фишера (**F_{n_1,n_2} **)**.

Стандартные распределения и их квантили

1. Стандартное нормальное распределение (Z)

Это распределение возникает как результат сложения многих независимых случайных воздействий.

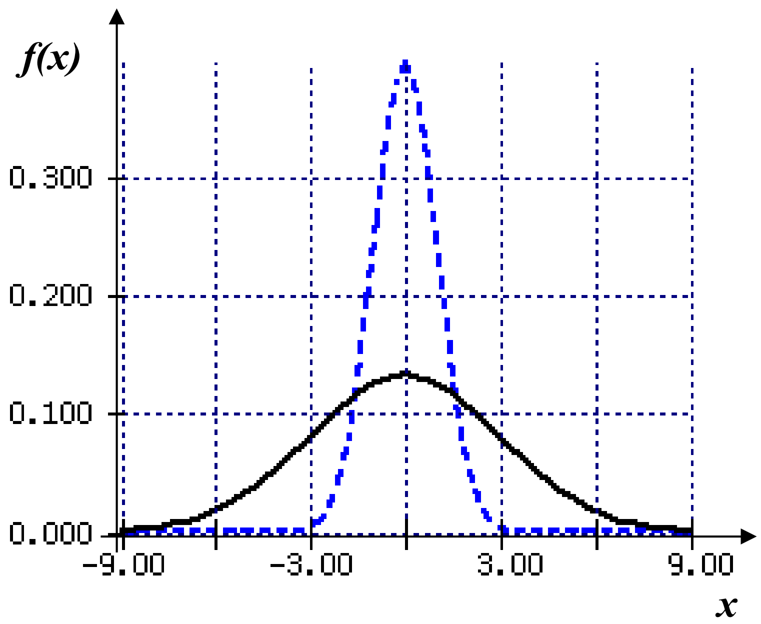


Стандартные распределения и их квантили

Нормальный закон определяется двумя параметрами:

μ - математическое ожидание;

σ - среднеквадратичное отклонение;

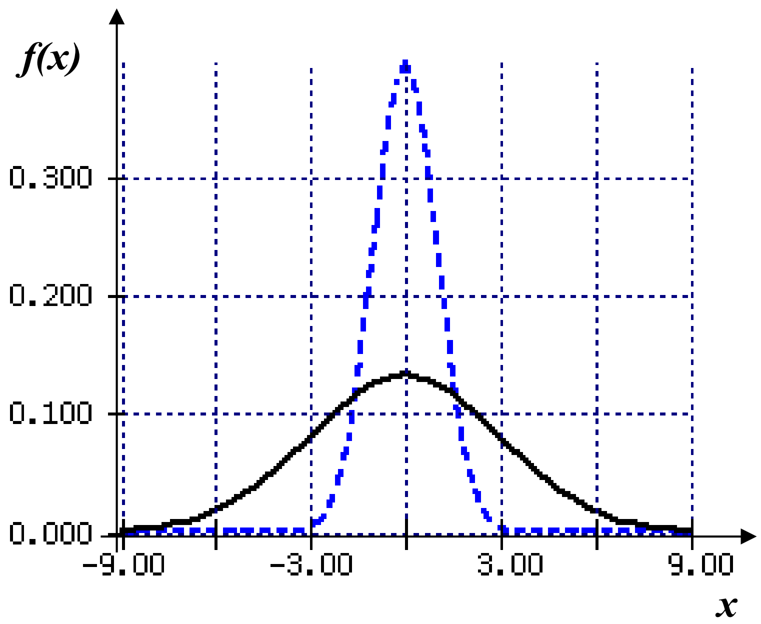


Стандартные распределения и их квантили

Нормальный закон определяется двумя параметрами:

μ - математическое ожидание;

σ - среднеквадратичное отклонение;



Обычно нормальное распределение используется в стандартном виде, где μ=**0**, σ=**1**. Переход от нормально распределенной величины x к величине со стандартным нормальным распределением z:

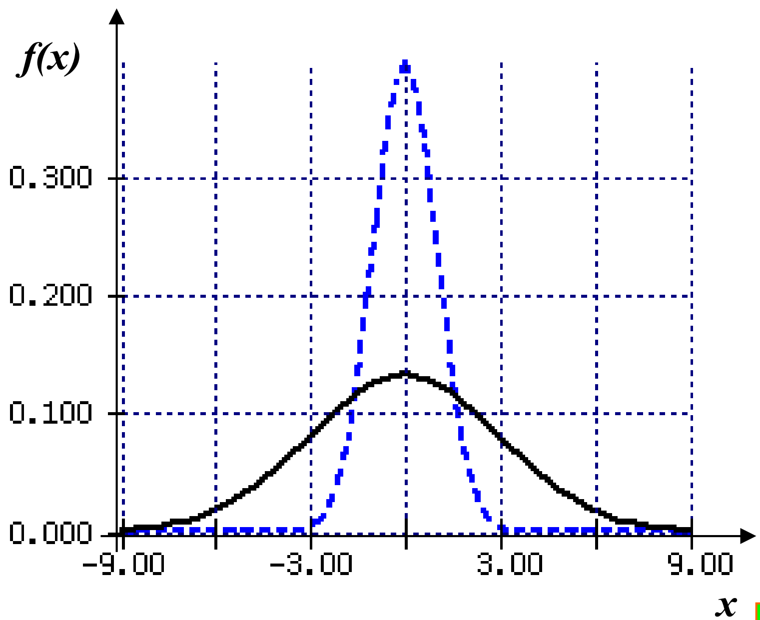
$$z = \frac{x - \mu_x}{\sigma_x};$$

Стандартные распределения и их квантили

Нормальный закон определяется двумя параметрами:

μ - математическое ожидание;

σ - среднеквадратичное отклонение;



Обычно нормальное распределение используется в стандартном виде, где ***μ*=0**, ***σ*=1**. Переход от нормально распределенной величины ***x*** к величине со стандартным нормальным распределением ***z***:

$$\boxed{z} = \frac{\boxed{x} - \mu_x}{\sigma_x};$$

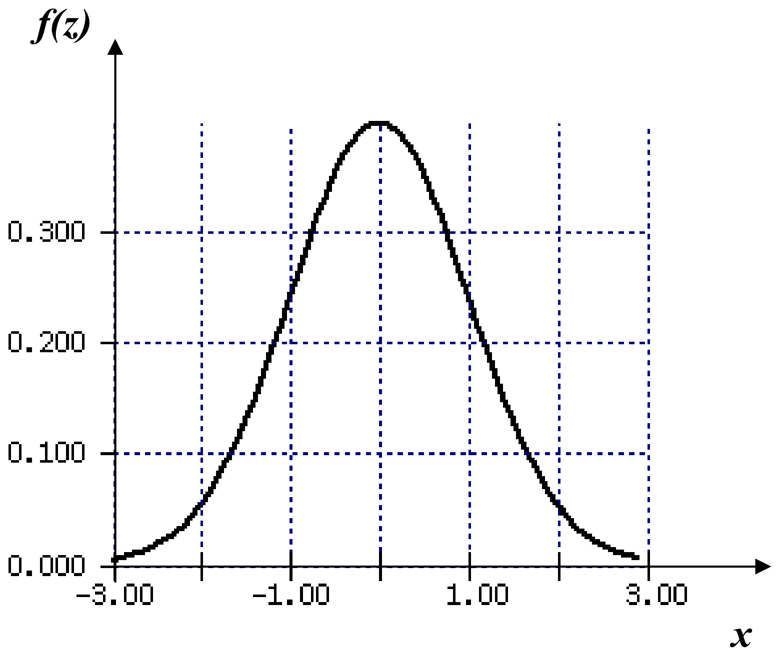
Где:

$\boxed{x}$ – нормально распределенная величина,

$\boxed{z}$ – величина со стандартным нормальным распределением.

Стандартные распределения и их квантили

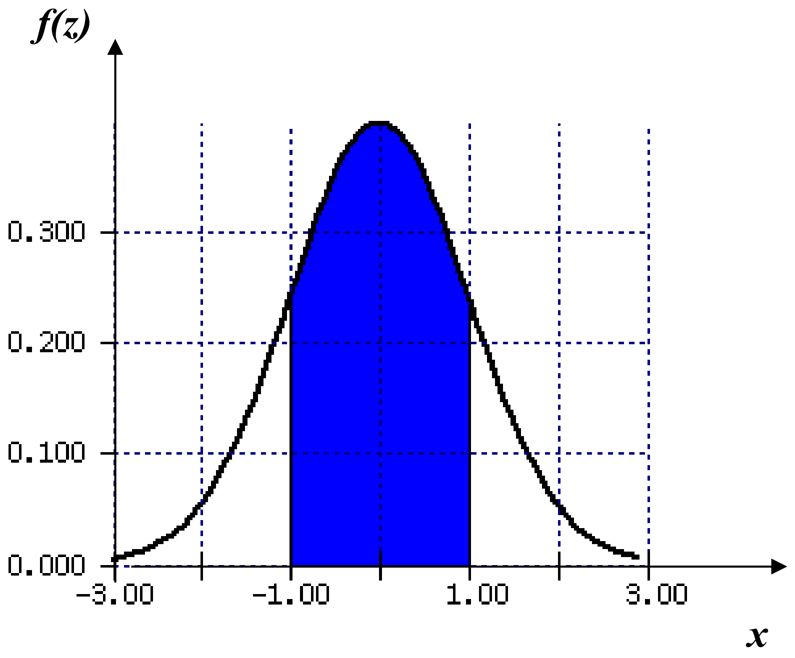
У нормального распределения есть три стандартных числа:



Стандартные распределения и их квантили

У нормального распределения есть три стандартных числа:

Вероятность попадания *x* в интервал [$\boldsymbol{\mu}$-$1\boldsymbol{\sigma}$; $\boldsymbol{\mu}$+$1\boldsymbol{\sigma}$] равна ≈ 68%.

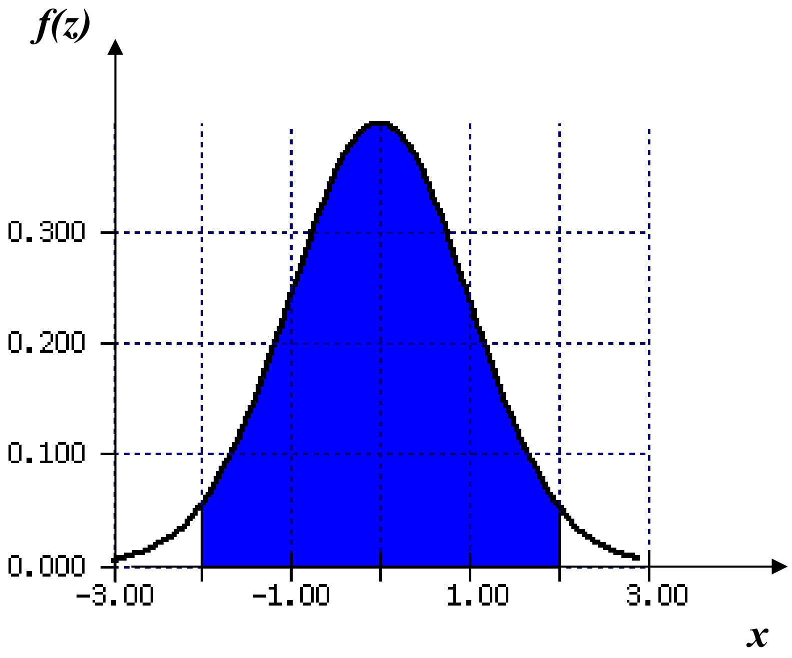


Стандартные распределения и их квантили

У нормального распределения есть три стандартных числа:

Вероятность попадания *x* в интервал [μ-1σ; μ+1σ] равна ≈ 68%.

Вероятность попадания *x* в интервал [μ-2σ; μ+2σ] равна ≈ 95%.



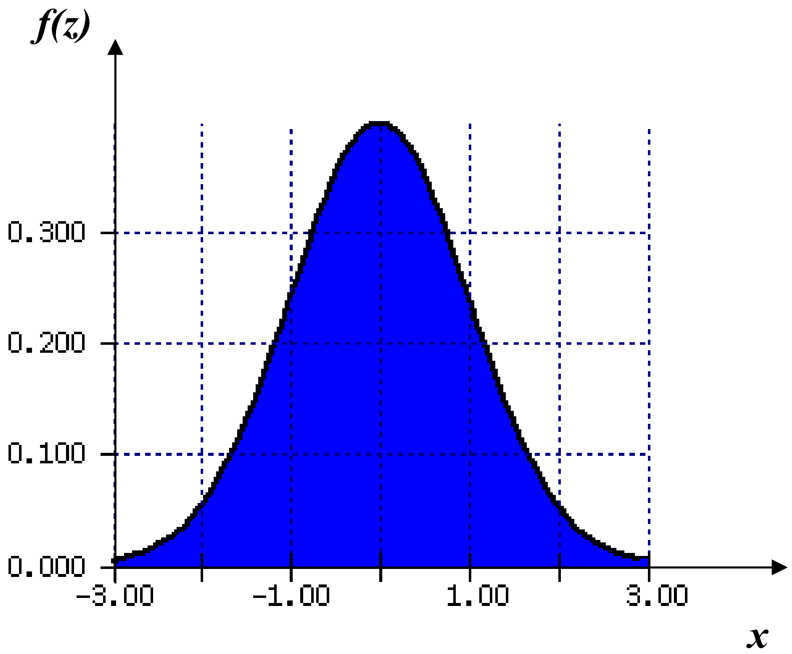
Стандартные распределения и их квантили

У нормального распределения есть три стандартных числа:

Вероятность попадания *x* в интервал [μ-1σ; μ+1σ] равна ≈ 68%.

Вероятность попадания *x* в интервал [μ-2σ; μ+2σ] равна ≈ 95%.

Вероятность попадания *x* в интервал [μ-3σ; μ+3σ] равна ≈ 99,7%.



Таким образом, на отрезке [-3σ, 3σ] находятся почти все значения. Это и есть так называемое **правило “трех сигм“**.

Стандартные распределения и их квантили

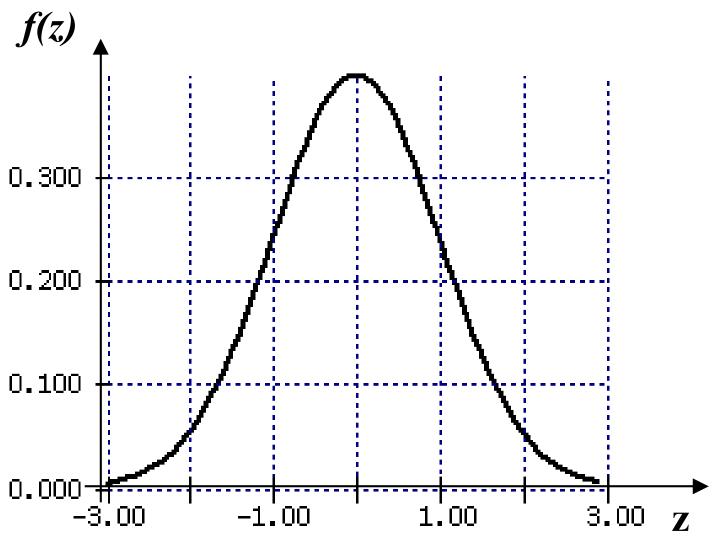
1. Интегральный закон распределения (z)

Функция ***f(x)*** показывает следующую важнейшую информацию: вероятность того, что величина ***x*** примет значение больше числа ***a*** и меньше числа ***b*** равна площади под кривой ***f(x)*** на отрезке **[*a*;*b*]**. Кроме того, площадь под всей кривой ***f(x)*** равна **1**.

Стандартные распределения и их квантили

1. Интегральный закон распределения (*z*)

Функция ***f(x)*** показывает следующую важнейшую информацию: вероятность того, что величина ***x*** примет значение больше числа ***a*** и меньше числа ***b*** равна площади под кривой ***f(x)*** на отрезке [***a***;***b***]. Кроме того, площадь под всей кривой ***f(x)*** равна **1**.



f(z) – плотность вероятности

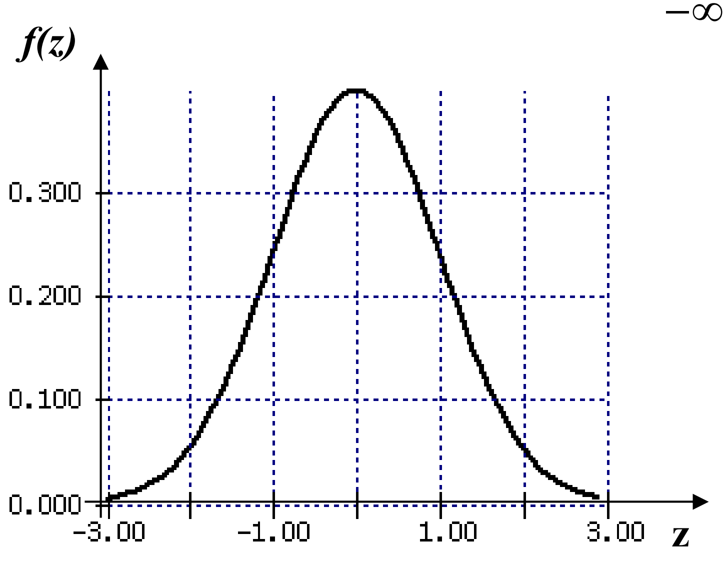
Стандартные распределения и их квантили

1. Интегральный закон распределения (*z*)

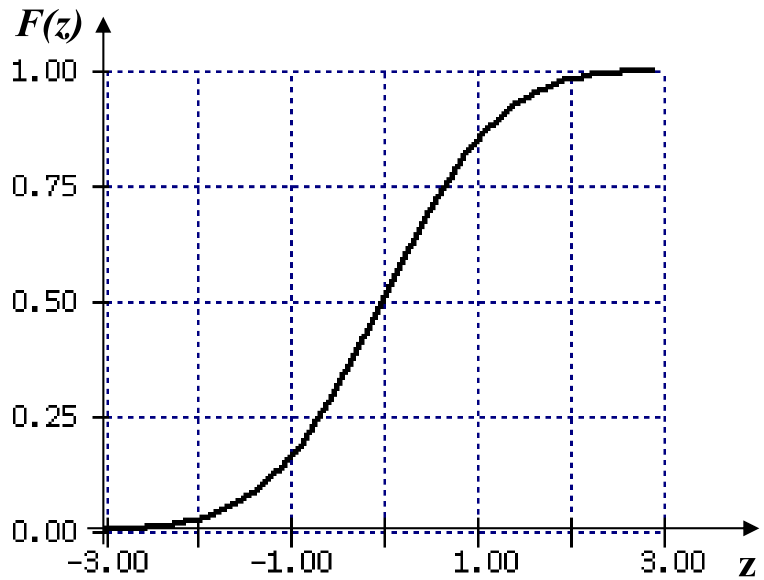
Функция ***f(x)*** показывает следующую важнейшую информацию: вероятность того, что величина ***x*** примет значение больше числа ***a*** и меньше числа ***b*** равна площади под кривой ***f(x)*** на отрезке **[*a*;*b*]**. Кроме того, площадь под всей кривой ***f(x)*** равна **1**.

Интегральная функция является интегралом от функции распределения

$$F(z) = \int_{-\infty}^{z} f(z)dz;$$

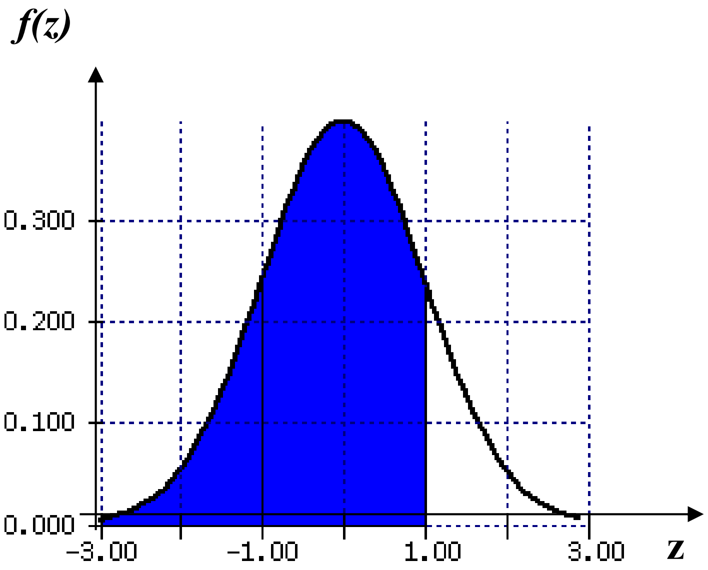


f(z) – плотность вероятности



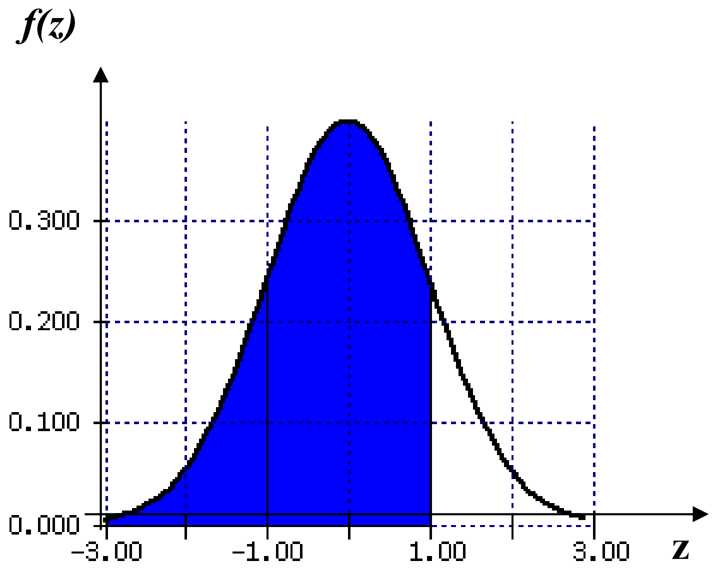
F(z) – интегральный закон распределения

1. Интегральный закон распределения (z)

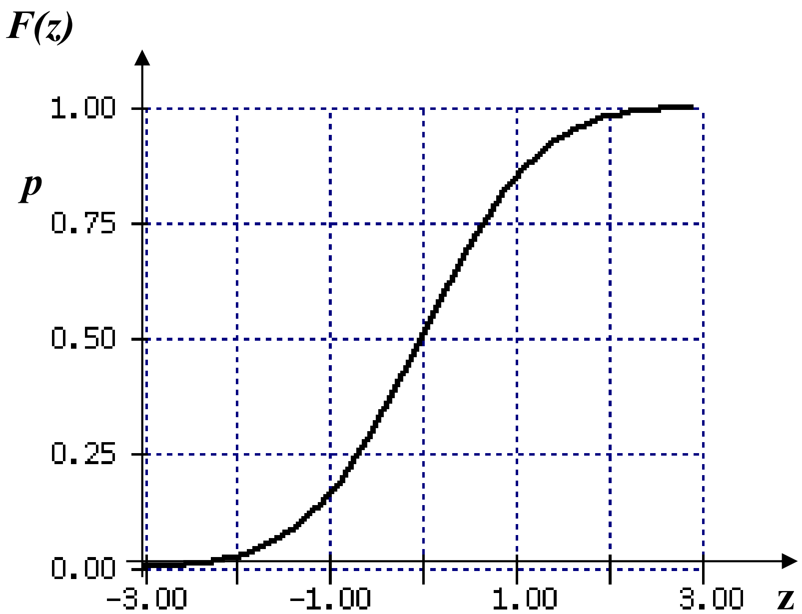


f(z) – плотность вероятности

1. Интегральный закон распределения (*z*)

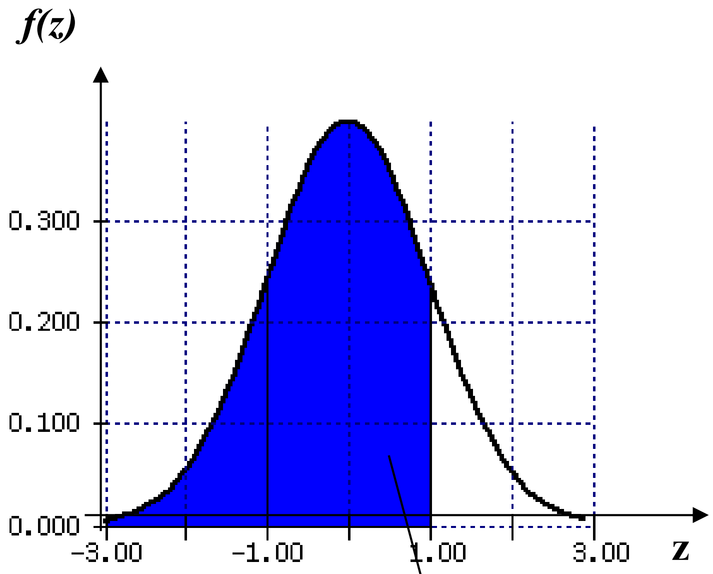


f(z) – плотность вероятности

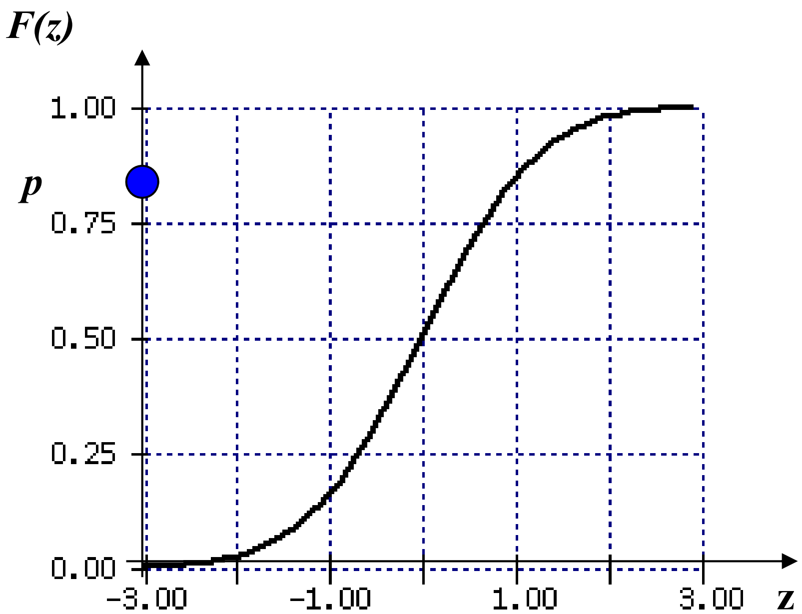


F(z) – интегральный закон распределения

1. Интегральный закон распределения (z)



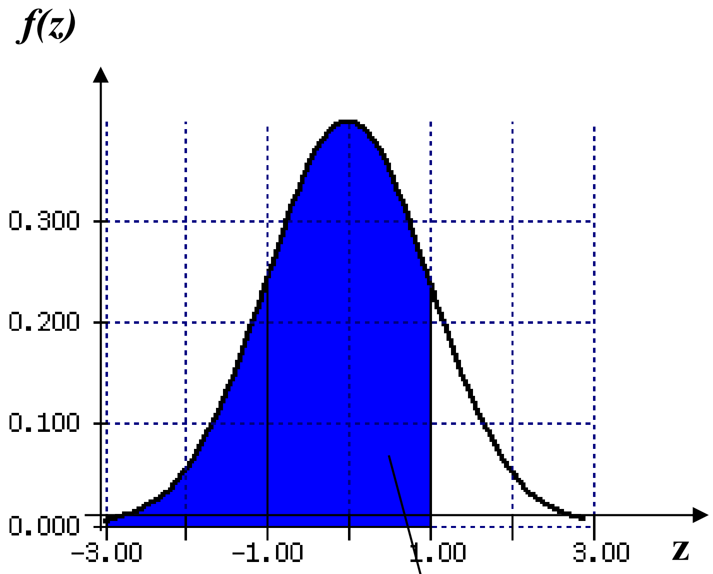
f(z) – плотность вероятности

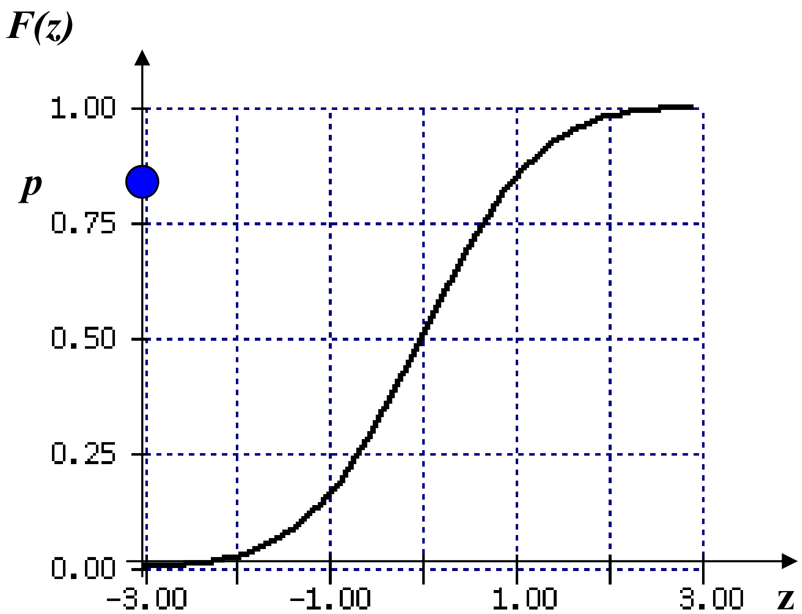


F(z) – интегральный закон распределения

эта площадь равна F(1), она дает точку на графике интегрального закона распределения

1. Интегральный закон распределения (z)





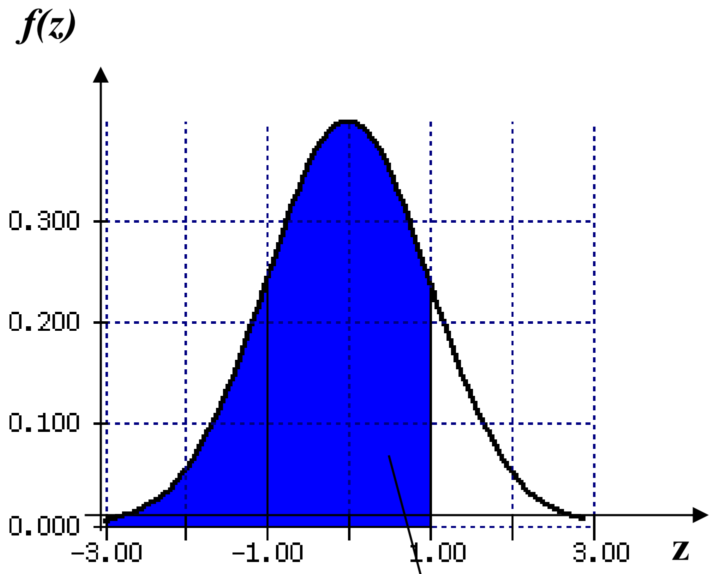
f(z) – плотность вероятности

F(z) – интегральный закон распределения

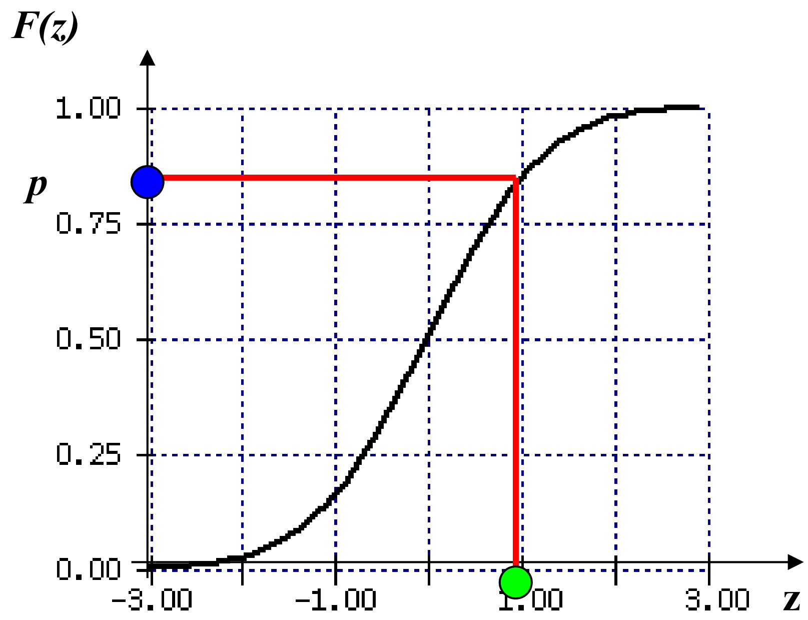
эта площадь равна F(1), она дает точку на графике интегрального закона распределения

Квантиль – это аргумент функции распределения, которой соответствует заданная вероятность ***p***.

1. Интегральный закон распределения (*z*)



f(z) – плотность вероятности



F(z) – интегральный закон распределения

эта площадь равна F(1), она дает точку на графике интегрального закона распределения

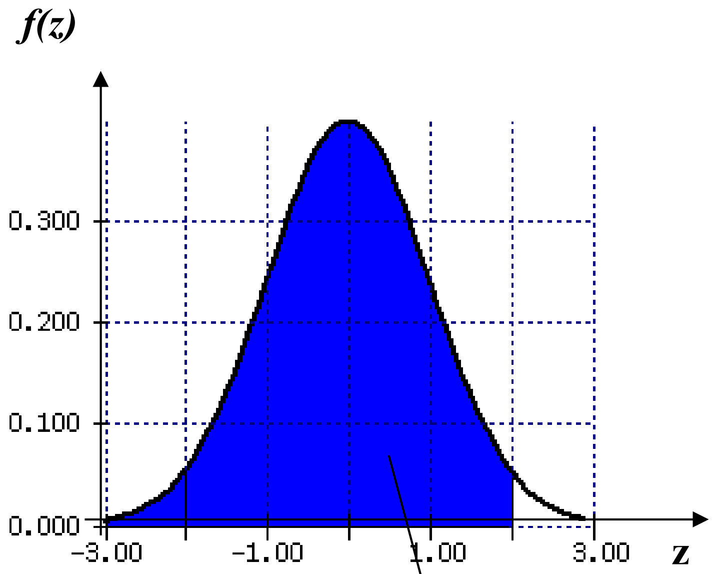
Квантиль – это аргумент функции распределения, которой соответствует заданная вероятность ***p***.

То есть, если выразить вероятность в виде $$p = F(z) = \int_{-\infty}^{z} f(z)dz;$$

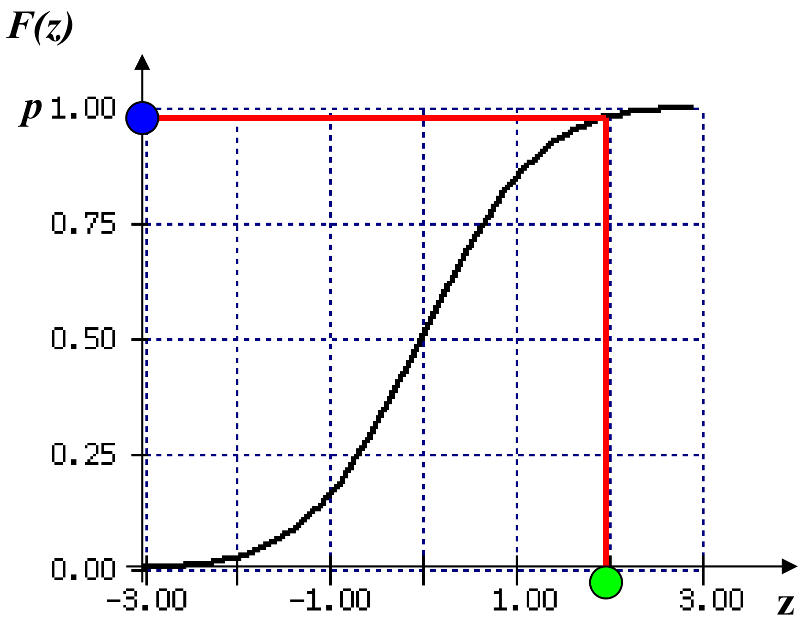
То квантилем будет ***z***, вычисленное из ***p***.

Стандартные распределения и их квантили

1. Интегральный закон распределения (*z*)



f(z) – плотность вероятности



F(z) – интегральный закон распределения

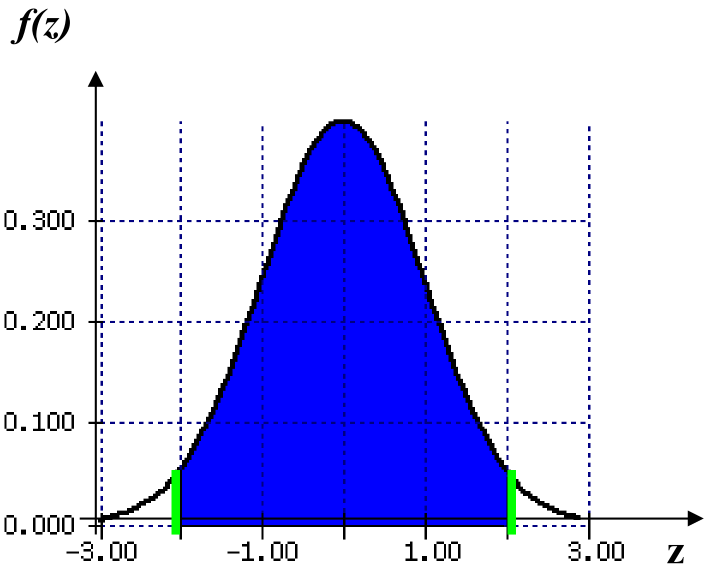
эта площадь равна F(2), она дает точку на графике интегрального закона распределения

Квантиль – это аргумент функции распределения, которой соответствует заданная вероятность ***p***.

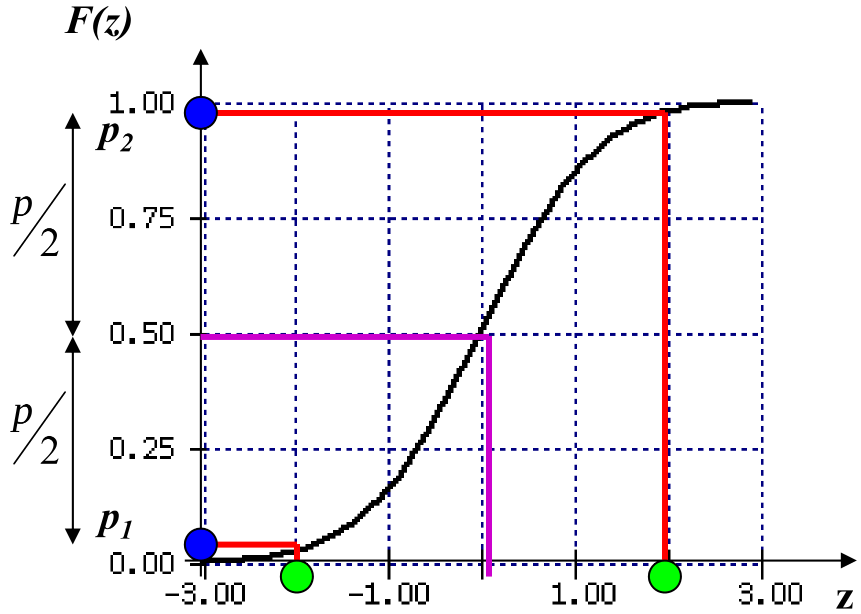
То есть, если выразить вероятность в виде $p = F(z) = \int_{-\infty}^{z} f(z)dz;$

То квантилем будет ***z***, вычисленное из ***p***.

Квантили нормального распределения (*z*)



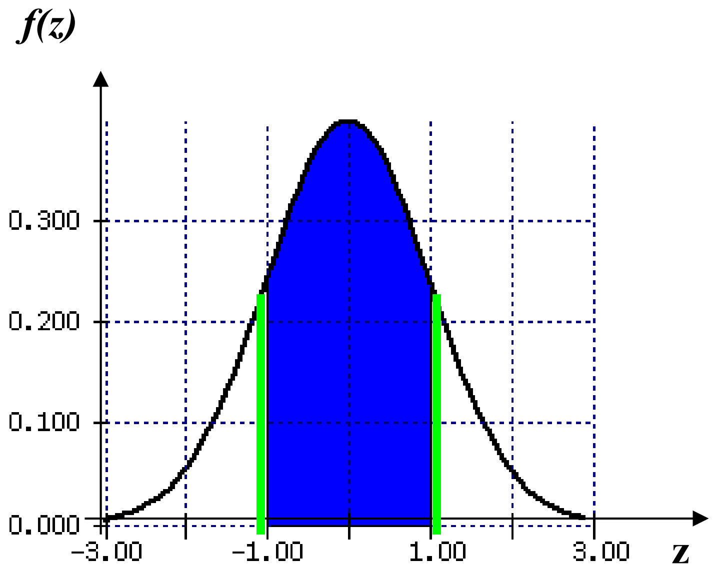
f(z) – плотность вероятности



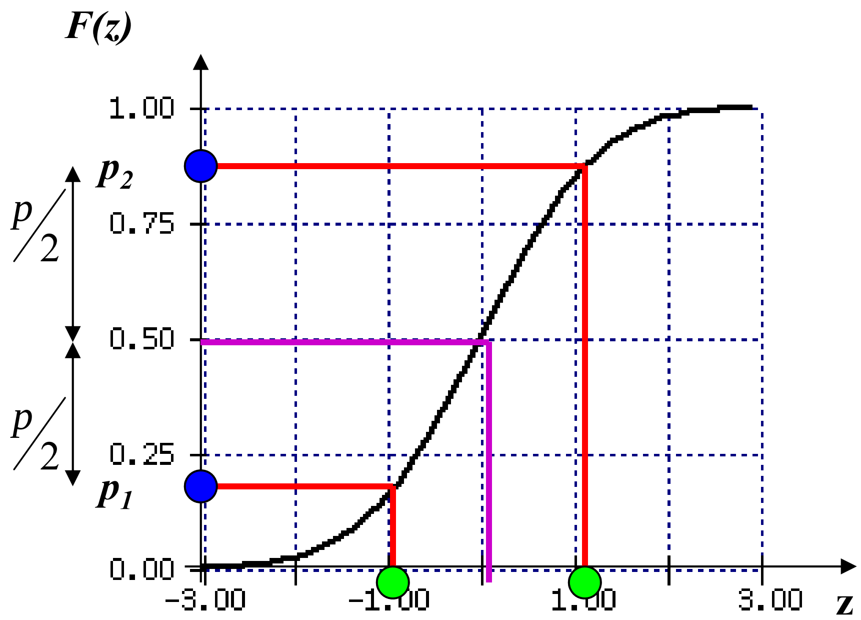
F(z) – интегральный закон распределения

Границам интервала на левом графике соответствуют значения p_1,p_2 на правом графике.

Квантили нормального распределения (*z*)



f(z) – плотность вероятности



F(z) – интегральный закон распределения

Границам интервала на левом графике соответствуют значения $\boldsymbol{p_1, p_2}$ на правом графике.

Стандартные распределения и их квантили

1. Таблица нормального распределения (z)

Так как формула нормального распределения очень сложна, используются статистические таблицы:

z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793

Стандартные распределения и их квантили

1. Таблица нормального распределения (z)

Так как формула нормального распределения очень сложна, используются статистические таблицы:

z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793

Чтобы определить квантиль по заданной вероятности, необходимо найти ближайшее к ней число в таблице и сложить значения соответствующих строки и столбца.

Строки соответствуют значениям z с точностью до десятой доли, а столбцы соответствуют их уточнениям до сотых долей.

Стандартные распределения и их квантили

1. Таблица нормального распределения (*z*)

Так как формула нормального распределения очень сложна, используются статистические таблицы:

z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793

Чтобы определить квантиль по заданной вероятности, необходимо найти ближайшее к ней число в таблице и сложить значения соответствующих строки и столбца.

Строки соответствуют значениям ***z*** с точностью до десятой доли, а столбцы соответствуют их уточнениям до сотых долей.

Например, известно, что ***z* = 0,31**, выделяем сотые доли, т.е. ***z* = 0,3+0,01**, значит ***F(z)*** находится на пересечении четвертой строки и второго столбца, и ***F(z)*** = **0,62172.**

Стандартные распределения и их квантили

1. Таблица нормального распределения (z)

В некоторых случаях таблицы бывают представлены в более компактном виде: остаются только дробные части всех или некоторых приведенных чисел.

То есть иногда в таблице отсутствуют некоторые нули и запятые.

z	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793

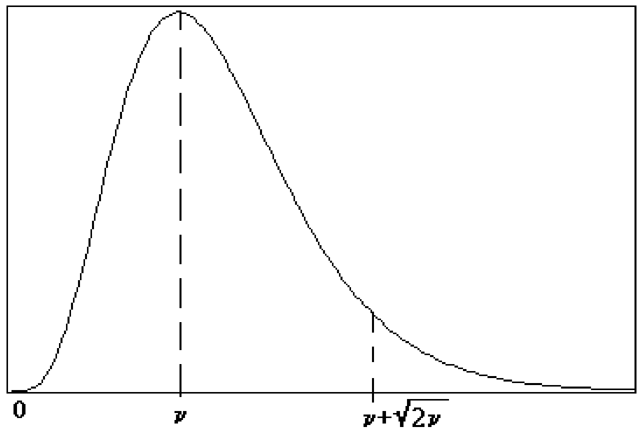
2. Распределение Пирсона или хи-квадрат (χ^2).

$$\chi_n^{\ 2} = z_1^{\ 2} + z_2^{\ 2} + \ldots + z_n^{\ 2};$$

Это распределение возникает как результат сложения квадратов нескольких величин, подчиняющихся нормальному закону с $\boldsymbol{\mu}$=**0**, $\boldsymbol{\sigma}$=**1**.

Число слагаемых ***n*** называется **числом степеней свободы**.

Смысл $\boldsymbol{f(\chi^2)}$ такой же, как и в нормальном законе: вероятность того, что величина $\boldsymbol{\chi^2}$ попадает в заданный интервал, равна площади под кривой $\boldsymbol{f(\chi^2)}$. Так, площадь под кривой на отрезке от **0** до $\boldsymbol{n+\sqrt{2n}}$ составляет более 90% всей площади под всей кривой $\boldsymbol{f(\chi^2)}$. Отсюда следует **правило “трех сигм“ для закона** $\boldsymbol{\chi^2}$: с вероятностью $\boldsymbol{p}$≥**0,9** случайная величина $\boldsymbol{\chi^2}$ не превосходит величины $\boldsymbol{n+\sqrt{2n}}$.



2. Таблица распределения хи-квадрат (χ^2).

n \ p	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995
1	0,00016	0,00098	0,0039	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0201	0,0506	0,1026	0,2107	4,61	5,99	7,38	9,21	10,60
3	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,34	12,84
4	0,297	0,484	0,711	1,064	7,78	9,49	11,14	13,28	14,86
5	0,554	0,831	1,15	1,61	9,24	11,07	12,83	15,09	16,75

Чтобы определить квантиль по заданной вероятности и числу степеней свободы, необходимо найти пересечение соответствующей строки и столбца.

Столбцам таблицы соответствуют вероятности, а строкам - число степеней свободы. В ячейках таблицы содержатся значения χ^2 (квантили).

2. Таблица распределения хи-квадрат (χ^2).

n \ p	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995
1	0,00016	0,00098	0,0039	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0201	0,0506	0,1026	0,2107	4,61	5,99	7,38	9,21	10,60
3	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,34	12,84
4	0,297	0,484	0,711	1,064	7,78	9,49	11,14	13,28	14,86
5	0,554	0,831	1,15	1,61	9,24	11,07	12,83	15,09	16,75

Чтобы определить квантиль по заданной вероятности и числу степеней свободы, необходимо найти пересечение соответствующей строки и столбца.

Столбцам таблицы соответствуют вероятности, а строкам - число степеней свободы. В ячейках таблицы содержатся значения χ^2 (квантили).

Например, для числа степеней свободы n=**3** и p =**0,975,** найдем χ^2 = **9,35**.

3. Распределение Стьюдента (t_n).

$$t_n = \frac{z}{\sqrt{\dfrac{\chi_n^{\ 2}}{n}}};$$

Это отношение стандартной нормальной величины к корню из хи-квадрат, деленной на число степеней свободы.

«Стьюдент» - это псевдоним английского статистика Уилльяма Госсета (William Gosset). **Пивная история Guinness. +доп инф.**

3. Таблица распределения Стьюдента (t_n).

n \ p	Односторонняя критическая область (p)							
	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
	Двусторонняя критическая область (p)							
	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,08	6,31	12,71	31,82	63,66	127,32	318,30	636,61
2	1,89	2,92	4,30	6,96	9,92	14,09	22,33	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	10,21	12,92
4	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
5	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87

Чтобы определить квантиль по заданной вероятности и числу степеней свободы, необходимо найти пересечение соответствующей строки и столбца.

Столбцам таблицы соответствуют вероятности, а строкам - число степеней свободы. В ячейках таблицы содержатся значения ***t*** (квантили).

3. Таблица распределения Стьюдента (t_n).

n \ p	Односторонняя критическая область (p)							
	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
	Двусторонняя критическая область (p)							
	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,08	6,31	12,71	31,82	63,66	127,32	318,30	636,61
2	1,89	2,92	4,30	6,96	9,92	14,09	22,33	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	10,21	12,92
4	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
5	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87

Чтобы определить квантиль по заданной вероятности и числу степеней свободы, необходимо найти пересечение соответствующей строки и столбца.

Столбцам таблицы соответствуют вероятности, а строкам - число степеней свободы. В ячейках таблицы содержатся значения ***t*** (квантили).

Например, мы ищем квантиль для односторонней критической области:

Для числа степеней свободы ***n*=4** и ***p* =0,025,** найдем ***t*** = **2,78**.

4. Распределение Фишера (F_{n_1, n_2}).

$$F_{n_1, n_2} = \frac{{\chi_{n_1}}^2 \Big/ n_1}{{\chi_{n_2}}^2 \Big/ n_2};$$

Это отношение двух хи-квадратов, деленных на число степеней свободы. Распределение имеет 2 степени свободы: для числителя и для знаменателя.

$\boldsymbol{n_1}$, $\boldsymbol{n_2}$ - число степеней свободы.

4. Распределение Фишера (F_{n_1,n_2}).

$$F_{n_1,n_2} = \frac{{\chi_{n_1}}^2 \big/ n_1}{{\chi_{n_2}}^2 \big/ n_2};$$

Это отношение двух хи-квадратов, деленных на число степеней свободы. Распределение имеет 2 степени свободы: для числителя и для знаменателя.

n_1, n_2 - число степеней свободы.

Обычно используется при сравнении двух дисперсий, так как дисперсия равна сумме квадратов отклонений от среднего значения, деленная на число точек.

4. Таблица распределения Фишера (F_{n_1, n_2}).

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74

Для распределения Фишера создано несколько отдельных таблиц, каждая из которых соответствует своему значению вероятности ***p***. На данном слайде изображена таблица для ***p*=0,95.**

4. Таблица распределения Фишера (F_{n_1, n_2}).

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74

Для распределения Фишера создано несколько отдельных таблиц, каждая из которых соответствует своему значению вероятности $\boldsymbol{p}$. На данном слайде изображена таблица для $\boldsymbol{p}$**=0,95.**

Строкам таблицы соответствуют значения $\boldsymbol{n_2}$, столбцам соответствуют значения $\boldsymbol{n_1}$.

4. Таблица распределения Фишера (F_{n_1,n_2}).

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74

Для распределения Фишера создано несколько отдельных таблиц, каждая из которых соответствует своему значению вероятности p. На данном слайде изображена таблица для **p=0,95.**

Строкам таблицы соответствуют значения n_2, столбцам соответствуют значения n_1.

Чтобы найти квантиль по заданной **вероятности** p и n_1, n_2, возьмем таблицу для соответствующей вероятности p и найдем значение на пересечении строки n_2 со столбцом n_1.

Например, для p=0,95, n_1=3, n_2=4, квантилем будет **6,59**.