ПРИЛОЖЕНИЕ 1

Все вероятности равны 50%. Либо случится, либо нет.
Мерфология, Логические предложения Кольварда

Типовые распределения

Johann Carl Friedrich Gauß

При проверке гипотез широкое применение находит ряд теоретических законов распределения. Наиболее важным из них является нормальное распределение. С ним связаны распределения хи-квадрат, Стьюдента, Фишера, а также интеграл вероятностей. Для указанных законов функции распределения аналитически в простой форме не представимы и определяются с использованием стандартных процедур прикладных программ.

Нормальное распределение

Распределение хи-квадрат

Распределение Стьюдента

Распределение Фишера

П1.1. Нормальное распределение

Нормальное распределение занимает важное место в статистике, поскольку очень многие эмпирические распределения биологических признаков приближаются к нему.

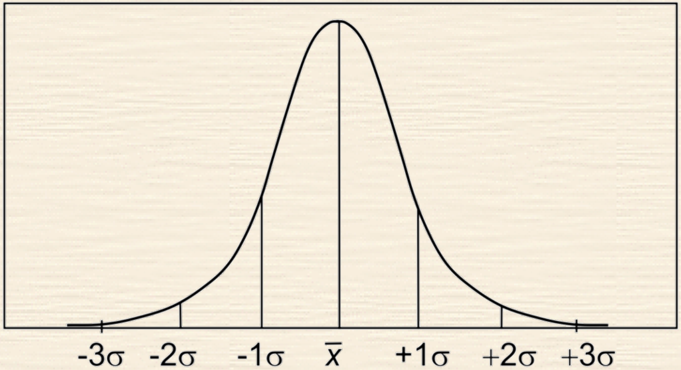

Рис. П1.1. Нормальная вариационная кривая

Размещение вариант в вариационном ряду при нормальном распределении характеризуется определенными закономерностями. В частности, в нормальной кривой отклонения от средней арифметической охватывают приблизи-тельно 6σ: 3σ справа от средней и 3σ слева.

Зная вариационную кривую распределения вариант по тому или иному признаку и предполагая, что распределение является нормальным, можно заранее предсказать, какой процент изученных особей (или вариант) укладывается в пределах $\pm 1\sigma$; в пределах $\pm 2\sigma$; в пределах $\pm 3\sigma$. А именно, в пределах

$\pm 1\sigma$ располагается 68,3% всех вариант данного ряда;

$\pm 2\sigma$ – 95,5% и в пределах $\pm 3\sigma$ – 99,7% всех вариант.

Например, при исследовании размеров раковин моллюсков рода *Benedictia* были получены следующие данные (мм): 32, 35, 37, 45, 41, 35, 39, 39, 45, 41. "Перенос" данных на генеральную совокупность позволяет предположить (рис. П1.2), что ≈ 68% популяции будут иметь средний размер раковины плюс-минус одно стандартное отклонение $\overline{x}\pm 1\sigma$, т.е. размеры раковин будут лежать в интервале от 34,62 до 43,18 мм.

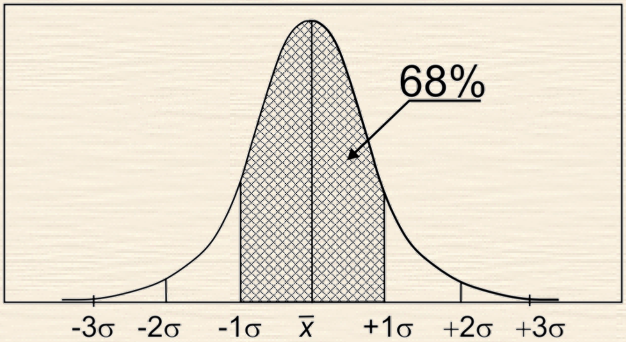

Рис. П1.2. Нормальное распределение значений размеров раковин у моллюсков *Benedictia* вокруг среднего

Около 98% популяции будет иметь размер раковины 38,9 плюс-минус два стандартных отклонения (8, 56), т.е. размер раковины будет лежать в интервале (30,34÷47,46) мм, и практически 100% будут лежать в интервале плюс-минус три стандартных отклонения от 38,9.

Нормальное распределение является наиболее важным в связи с центральной предельной теоремой теории вероятностей: распределение суммы независимых случайных величин стремится к нормальному с увеличением их количества при произвольном законе распределения отдельных слагаемых, если слагаемые обладают конечной дисперсией.

Так как реальные физические, химические, биологические явления часто представляют собой результат суммарного воздействия многих факторов, то в таких случаях нормальное распределение является хорошим приближением наблюдаемых значений. Функция плотности нормального распределения

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\overline{x})^2}{2\sigma^2}\right)$$

– унимодальная, симметричная, аргумент x может принимать любые действительные значения (см. рис. П1.3).

Стандартизирующее преобразование. Если случайная величина X имеет известное математическое ожидание μ и дисперсию σ^2 то случайная величина Y=X-μ называется центрированной, величина Y= X/σ – нормированной, а $Y=(X$-$\mu)/\sigma$ – стандартизированной. Последнее преобразование называется стандартизирующим и используется на практике для получения стандартизированных выборок (в качестве значений μ и σ обычно берутся выборочные среднее и стандартное отклонение). Для выполнения этого преобразования в Excel есть специальная функция НОРМАЛИЗАЦИЯ(x; $\overline{x}$; σ) {STANDARDIZE (x; $\overline{x}$; σ)}.

Отметим, что стандартизирующее преобразование не изменяет тип распределения, а изменяет только значения математического ожидания и дисперсии.

Функция плотности нормального распределения стандартизованной величины u имеет вид $f(u) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{u^2}{2}\right)$.

ПЛОТНОСТЬ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

x	P(x,σ=1)	P(x,σ=1,5)
-3,00	0,004	0,036
-2,75	0,009	0,050
-2,50	0,018	0,066
-2,25	0,032	0,086
-2,00	0,054	0,109
-1,75	0,086	0,135
-1,50	0,130	0,161
-1,25	0,183	0,188
-1,00	0,242	0,213
-0,75	0,301	0,235
-0,50	0,352	0,252
-0,25	0,387	0,262
0,00	0,399	0,266
0,25	0,387	0,262
0,50	0,352	0,252
0,75	0,301	0,235
1,00	0,242	0,213
1,25	0,183	0,188
1,50	0,130	0,161
1,75	0,086	0,135
2,00	0,054	0,109
2,25	0,032	0,086
2,50	0,018	0,066
2,75	0,009	0,050
3,00	0,004	0,036

=1/КОРЕНЬ(2*ПИ())/F$13/EXP(((A4-F$11)/F$13)^2/2)

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\overline{x})^2}{2\sigma^2}\right)$$

$\overline{x}$ =	0,00
σ_1 =	1,00
σ_2 =	1,50

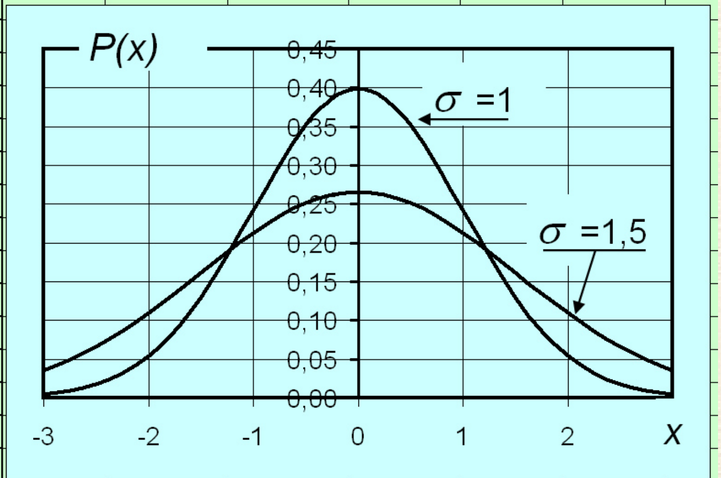

Рис. П1.3. Плотность вероятности нормального распределения

П1.2. Распределение χ^2 хи-квадрат

Распределению хи-квадрат (χ^2-распределение) с k степенями свободы соответствует распределение суммы $\chi^2 = \sum u_i^2$ квадратов n стандартизованных случайных величин u_i, каждая из которых распределена по нормальному закону, причем k из них независимы, $n > k$. Функция плотности распределения χ^2 с k степенями свободы

$$P(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\left(\frac{k}{2}-1\right)} exp\left(-\frac{x}{2}\right),$$ где $x = \chi^2$, $\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx$ – гамма-функция.

Число степеней свободы k определяет количество независимых слагаемых в выражении для χ^2. Функция плотности при k, равном одному или двум, – монотонная, а при k>2 – унимодальная, несимметричная, рис. П1.4.

Математическое ожидание и дисперсия величины χ^2 равны соответственно k и $2k$. Распределение χ^2 является частным случаем более общего гамма-распределения, а величина, равная корню квадратному из хи-квадрат с двумя степенями свободы, подчиняется распределению Рэлея.

С увеличением числа степеней свободы (k>30) распределение χ^2 приближается к нормальному распределению с математическим ожиданием k и дисперсией $2k$. В таких случаях критическое значение $\chi^2(k;\alpha) >> u_{1-\alpha}(k,2k)$, где $u_{1-\alpha}(k,2k)$ – квантиль нормального распределения. Погрешность аппроксимации не превышает нескольких процентов.

Плотность распределения хи-квадрат

$$P(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\left(\frac{k}{2}-1\right)} exp\left(-\frac{x}{2}\right)$$

x	*P(k=3)*	*P(k=6)*	*P(k=10)*
0,00	0,0000	0,0000	0,0000
1,00	0,2420	0,0379	0,0008
1,50	0,2308	0,0664	0,0031
2,00	0,2076	0,0920	0,0077
2,50	0,1807	0,1119	0,0146
3,00	0,1542	0,1255	0,0235
3,50	0,1297	0,1330	0,0340
4,00	0,1080	0,1353	0,0451
4,50	0,0892	0,1334	0,0563
5,00	0,0732	0,1283	0,0668
5,50	0,0598	0,1209	0,0762
6,00	0,0487	0,1120	0,0840
6,50	0,0394	0,1024	0,0901
7,00	0,0319	0,0925	0,0944
7,50	0,0257	0,0827	0,0969
8,00	0,0207	0,0733	0,0977
8,50	0,0166	0,0644	0,0970
9,00	0,0133	0,0562	0,0949
9,50	0,0106	0,0488	0,0918
10,00	0,0085	0,0421	0,0877
10,50	0,0068	0,0362	0,0831
11,00	0,0054	0,0309	0,0779
11,50	0,0043	0,0263	0,0725
12,00	0,0034	0,0223	0,0669
12,50	0,0027	0,0189	0,0614

=H$7*A4^H$8/EXP(A4/2)

=1/EXP(ГАММАНЛОГ(H6/2))/2^(H6/2)

k=	3	6	10
c=	0,40	0,0625	0,0013
p=	0,5	2	4

=H6/2-1

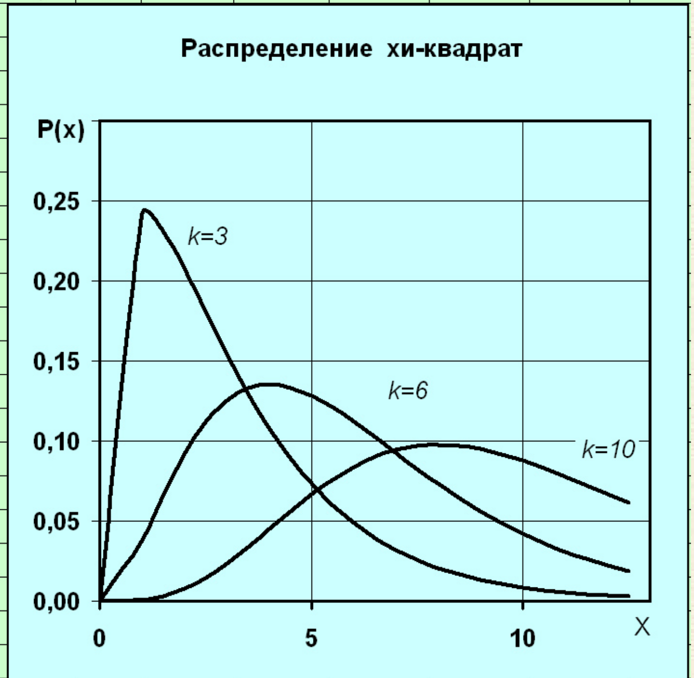

Рис. П1.4. Распределения плотности вероятности χ^2

П1.3. Распределение Стьюдента

Распределение Стьюдента (t-распределение, предложенное в 1908 г. английским статистиком В. Госсетом, публиковавшим научные труды под псевдонимом *Student*) характеризует распределение случайной величины $t = \frac{u_0}{\sqrt{(u_1^2+u_2^2+\cdots+u_k^2)/k}}$, где $u_0, u_1, \ldots, u_k$ взаимно независимые нормально распределенные случайные величины с нулевым средним и конечной дисперсией.

Аргумент t не зависит от дисперсии слагаемых. Функция плотности распределения Стьюдента

$$P(t) = \frac{1}{\sqrt{\pi k}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left[1 + \frac{t^2}{k}\right]^{-\frac{k+1}{2}}.$$

Величина k характеризует количество степеней свободы. Плотность распределения – унимодальная и симметричная функция, похожая на нормальное распределение, рис. П1.5.

Область изменения аргумента t от $-\infty$ до $+\infty$. Математическое ожидание и дисперсия равны (при k>2) 0 и $k/(k-2)$ соответственно. По сравнению с нормальным распределение Стьюдента более пологое, оно имеет меньшую дисперсию. Это отличие заметно при небольших значениях k, что следует учитывать при проверке статистических гипотез (критические значения аргумента распределения Стьюдента превышают аналогичные показатели нормального распределения).

ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА

$$P(t) = \frac{1}{\sqrt{\pi k}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left[1 + \frac{t^2}{k}\right]^{-\frac{k+1}{2}}$$

x	P (k=3)	P (k=30)
-3,00	0,023	0,007
-2,75	0,030	0,012
-2,50	0,039	0,021
-2,25	0,051	0,035
-2,00	0,068	0,057
-1,75	0,090	0,088
-1,50	0,120	0,129
-1,25	0,159	0,180
-1,00	0,207	0,238
-0,75	0,261	0,297
-0,50	0,313	0,348
-0,25	0,353	0,383
0,00	0,368	0,396
0,25	0,353	0,383
0,50	0,313	0,348
0,75	0,261	0,297
1,00	0,207	0,238
1,25	0,159	0,180
1,50	0,120	0,129
1,75	0,090	0,088
2,00	0,068	0,057
2,25	0,051	0,035
2,50	0,039	0,021
2,75	0,030	0,012
3,00	0,023	0,007

=F$8*(1+A4*A4/F$7)^F$9

	E	F
k=	3	30
c=	0,37	0,40
p=	-2,00	-15,50

=EXP(ГАММАНЛОГ((F7+1)/2)
-ГАММАНЛОГ(F7/2))/КОРЕНЬ(ПИ()*F7)

=-(F7+1)/2

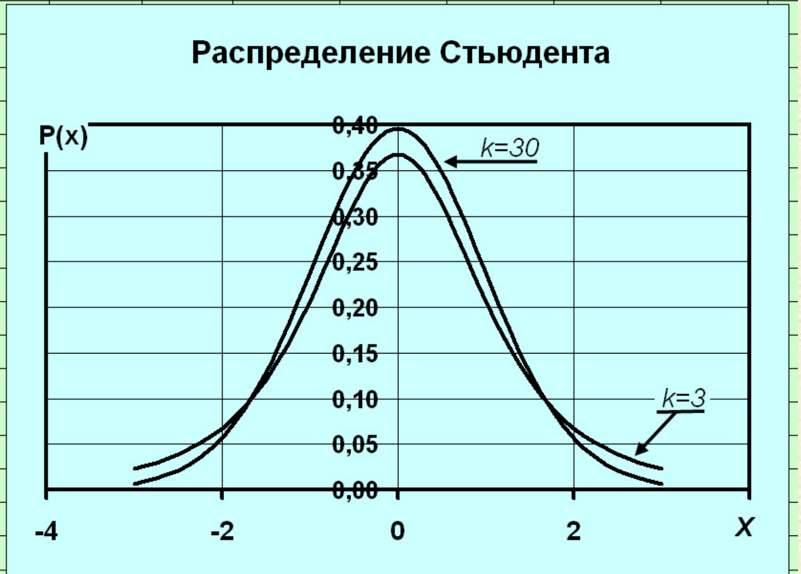

Рис. П1.5. Распределения плотности вероятности Стьюдента

Таблицы распределения (и возвращаемые значения функций электронных таблиц) содержат значения для односторонней $\int_{t(k;\alpha)}^{\infty} f(t)dt = \alpha$ или двусторонней $\int_{-t(k;\alpha)}^{t(k;\alpha)} f(t)dt = \alpha$ критической области.

Распределение Стьюдента применяется для описания ошибок выборки при k<30. При k >100 данное распределение практически соответствует нормальному, для 30<k<100 различия между распределением Стьюдента и нормальным распределением составляют несколько процентов. Поэтому относительно оценки ошибок малыми считаются выборки объемом не более 30 единиц, большими – объемом более 100 единиц. При аппроксимации распределения Стьюдента нормальным распределением для односторонней критической области вероятность

$$P\{t > t(k;\alpha)\} = u_{1-\alpha}\left(0, \frac{k}{k-2}\right),$$

где $u_{1-\alpha}\left(0, \frac{k}{k-2}\right)$ – квантиль нормального распределения. Аналогичное соотношение можно составить и для двусторонней критической области.

П1.4. Распределение Фишера

Распределению Фишера (F-распределению Фишера – Снедекора) подчиняется случайная величина $x = \frac{y_1/k_1}{y_2/k_2}$, равная отношению двух случайных величин y_1 и y_2, имеющих хи-квадрат распределение с k_1 и k_2 степенями свободы. Область изменения аргумента x от 0 до +∞. Плотность распределения

$$P(x) = \left(\frac{k_1}{k_2}\right)^{k_1/2} \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} x^{(k_1-2)/2} \left(1+\frac{k_1}{k_2}x\right)^{-(k_1+k_2)/2}.$$

В этом выражении k_1 обозначает число степеней свободы величины y_1 с большей дисперсией, k_2 – число степеней свободы величины y_2 с меньшей дисперсией. Плотность распределения – унимодальная, несимметричная (см. рис. П1.6).

Аналогами используемых в приведенных построениях функций русифицированных сред являются следующие английские.

квадратный корень	КОРЕНЬ(...)	SQRT(...)
число π	ПИ( )	PI( )
натуральный логарифм гамма-функции	ГАММАНЛОГ(...)	GAMMALN(...)

Плотность вероятности распределения Фишера

x	P (k2=10)	P (k2=30)
0,00	0,000	0,000
0,10	0,001	0,000
0,20	0,034	0,008
0,30	0,175	0,070
0,40	0,398	0,242
0,50	0,611	0,500
0,60	0,755	0,762
0,70	0,819	0,953
0,80	0,820	1,044
0,90	0,779	1,039
1,00	0,714	0,967
1,10	0,640	0,855
1,20	0,564	0,729
1,30	0,492	0,603
1,40	0,426	0,489
1,50	0,367	0,390
1,60	0,316	0,307
1,70	0,271	0,239
1,80	0,233	0,186
1,90	0,200	0,143
2,00	0,172	0,110
2,10	0,148	0,084
2,20	0,128	0,065
2,30	0,111	0,050
2,40	0,096	0,038

=F$11*$A4^F$12*(1+F$9/F$10*$A4)^F$13

$$P(x) = \left(\frac{k_1}{k_2}\right)^{k_1/2} \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} x^{(k_1-2)/2} \left(1+\frac{k_1}{k_2}x\right)^{-(k_1+k_2)/2}$$

k1 =	20	20
k2 =	10	30
c=	1,03E+07	3,40E+05
p1=	9	9
p2=	-15	-25

=EXP(ГАММАНЛОГ(F9/2+F10/2)
-ГАММАНЛОГ(F9/2)
-ГАММАНЛОГ(F10/2))*(F9/F10)^(F9/2)

=(F9-2)/2

=-(F9+F10)/2

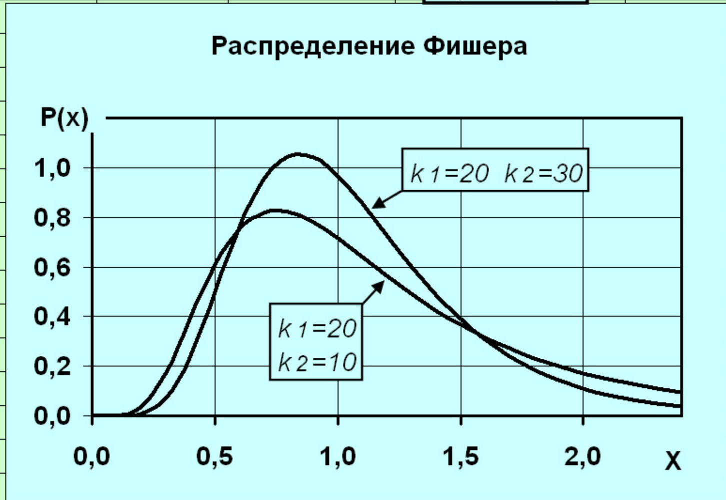

Рис. П1.6. Распределения плотности вероятности Фишера