СТАТИСТИКА.

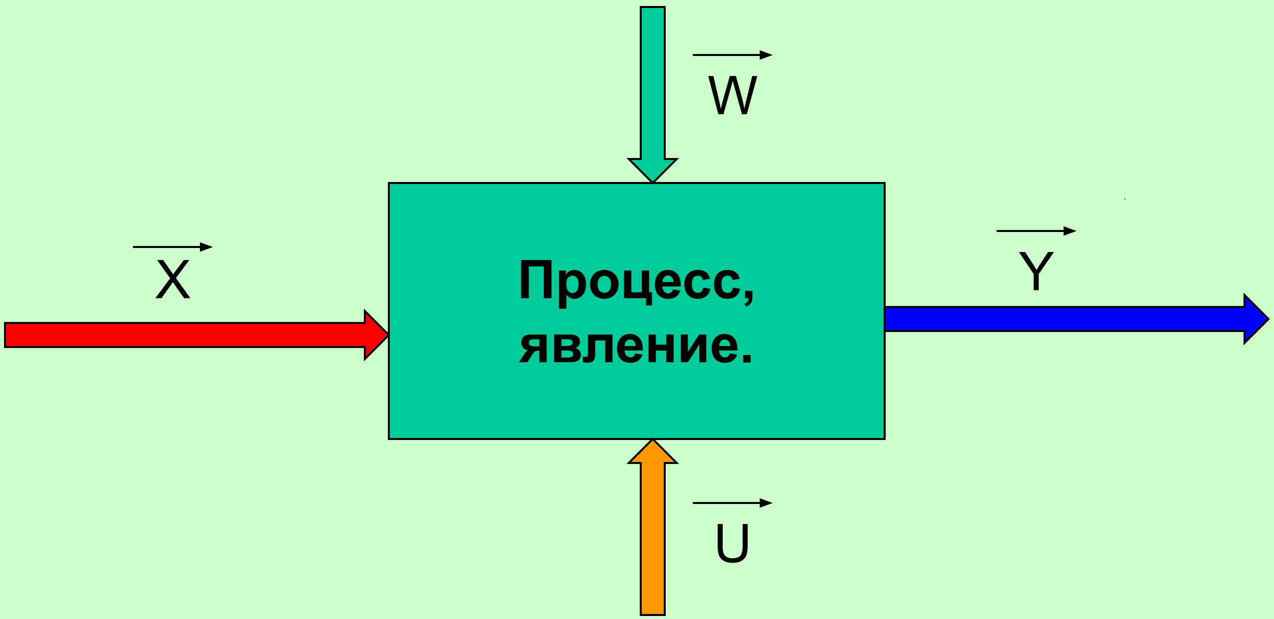
Аналитическая статистика.

Лекция 1. Теоретические распределения в анализе вариационных рядов.

Автор: Равичев Л.В.
РХТУ им. Д.И.Менделеева
Кафедра управления технологическими инновациями
Москва - 2013

Общие сведения о математическом моделировании

Различают два вида зависимостей между явлениями и процессами: ***функциональную*** и ***стохастическую*** (вероятностную, статистическую).



Моделирование рядов распределения

Если имеется эмпирический ряд распределения, то необходимо найти функцию распределения, т.е. подобрать такую теоретическую кривую распределения, которая наиболее полно отображала бы закономерность распределения. Нахождение функции кривой распределения называется ***моделированием эмпирического ряда распределения.***

Для аппроксимации (выравнивания) эмпирических кривых распределения и сопоставления их с теоретическими в статистике часто используются ***нормальное распределение:***

$$\varphi_{(t)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

где $\phi_{(t)}$ – ордината кривой нормального распределения; $t = (x - \bar{x})/\sigma$ - стандартное отклонение; x – варианты ряда; $\bar{x}$ – средняя величина вариант; σ - стандарт.

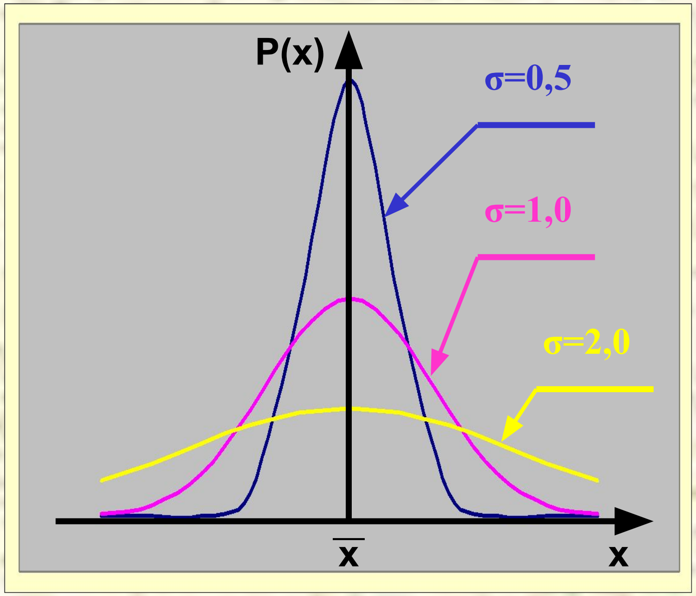
Моделирование рядов распределения

Основные свойства кривой нормального распределения:

- $\phi_{(t)}$ - функция нормального распределения – четная, т.е. $\phi_{(-t)} = \phi_{(+t)}$;
- функция имеет бесконечно малые значения при $t = \pm\infty$;
- функция имеет максимум при $t = 0$;
- при $t = \pm1$ функция имеет точки перегиба;

В статистике часто используют функцию ***плотности распределения:***

$$P_{(x)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\overline{X})^2}{2\sigma^2}}$$



Моделирование рядов распределения

Связь между теоретической ***нормированной*** функцией нормального распределения и теоретической ***денормированной*** функцией нормального распределения для интервального вариационного ряда определяется соотношением:

$$\boxed{\varphi_{m(t_i)} = A \cdot \varphi_{(t_i)}}$$

где А – коэффициент нормировки, который для распределения с равными интервалами $\Delta x = k$ рассчитывается с помощью соотношения:

$$\boxed{A = \frac{k\sum_{i=1}^{n} f_i}{\sigma}}$$

f_i - частота i-го интервала ряда.

Расчет теоретических частот нормального распределения

<u>Пример.</u> В приведенной таблице показано распределение ткачих по степени выполнения норм выработки. Исходя из предположения о нормальном законе распределения определить теоретические частоты.

Группы ткачих по степени выполнения норм, % x	Число ткачих f	Середина интервала x'
До 100	2	95
100-110	15	105
110-120	20	115
120-130	32	125
130-140	18	135
140-150	9	145
Свыше 150	4	155
Итого	100	-

1. Находим среднее значение выполнения норм по формуле:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n}{f_1 + f_2 + \ldots + f_n}$$

Среднее значение выполнения норм $\bar{x}$ =124,20%,

2. Находим взвешенное квадратическое отклонение (стандарт) по формуле:

$$\sigma = \pm \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{n} f_i}}$$

Стандарт σ= 13,69%.

Расчет теоретических частот нормального распределения

3. Находим значения параметра t.

4. Находим значения параметра t^2.

$$A = \frac{k\sum_{i=1}^{n} f_i}{\sigma} = \frac{10 \cdot 100}{13{,}69} = 73{,}046$$

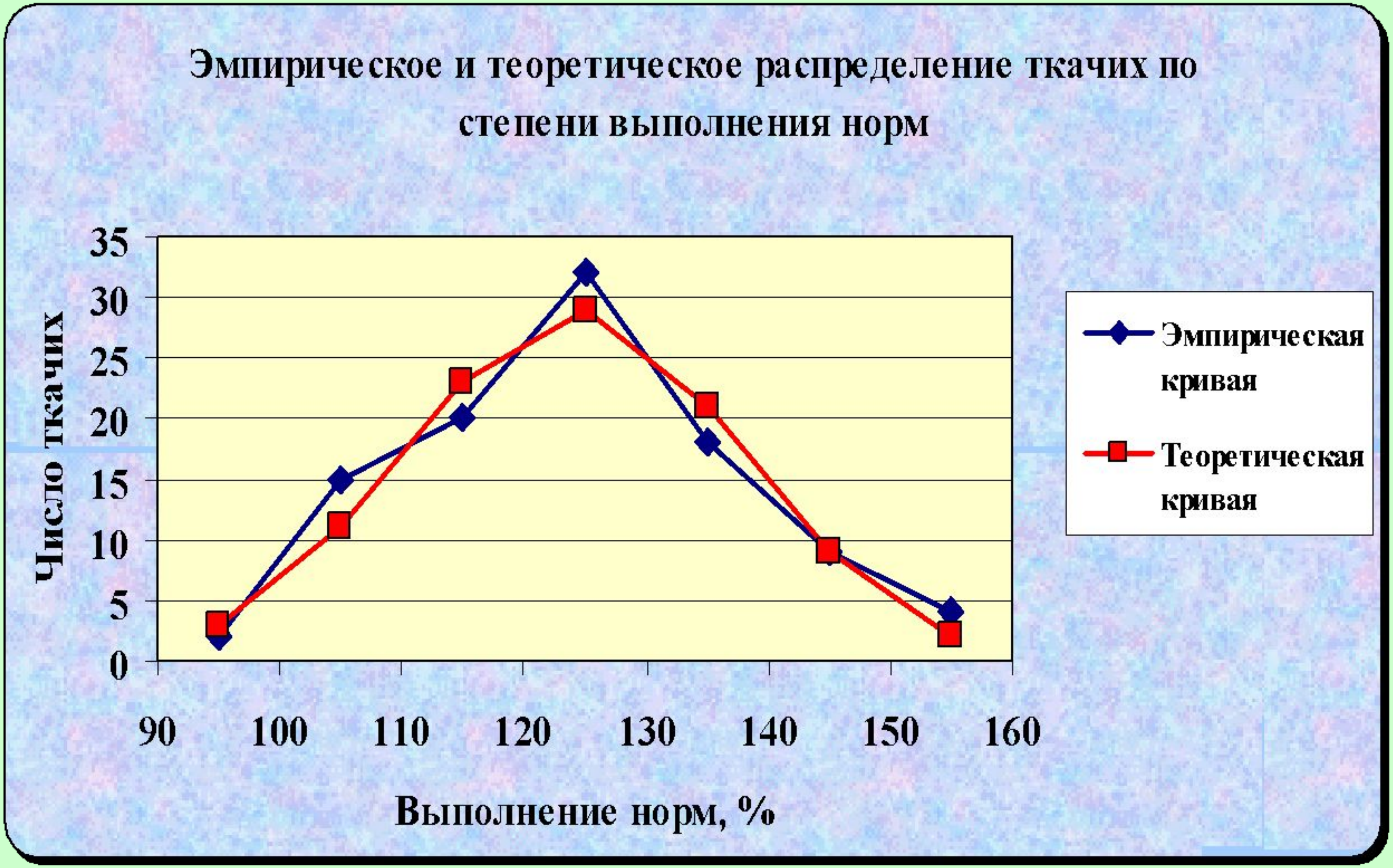
5. Находим значения теоретической нормированной функции $\phi_{(t)}$.

6. Находим значение коэффициента А.

7. Находим теоретические частоты $\phi_{m(t)}$ и f_m.

Группы ткачих по степени выполнения норм, % x	Число ткачих f	Середина интервала x'	$t=(x'-\bar{x})/\sigma$	t^2	Теоретические частоты		
					$\phi_{(t)}$	$\phi_{m(t)}$	f_m
До 100	2	95	-2,133	4,551	0,04099	2,99	3
100-110	15	105	-1,403	1,968	0,14916	10,90	11
110-120	20	115	-0,672	0,452	0,31828	23,25	23
120-130	32	125	0,058	0,003	0,39826	29,10	29
130-140	18	135	0,789	0,623	0,29223	21,35	21
140-150	9	145	1,520	2,309	0,12574	9,19	9
Свыше 150	4	155	2,250	5,063	0,03173	2,32	2
Итого	100	-	-	-	-	99,10	99

Расчет теоретических частот нормального распределения



Методы расчета значений теоретической нормированной функции $\varphi_{(t)}$

1. С помощью таблицы значений нормированной функции:

Целые и десятые доли t	Сотые доли t						
	0	1	2	3	…	8	9
0,0	0,3989	0,3989	0,3989	0,3988	…	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	…	0,3925	0,3918
…	…	…	…	…	…	…	…
2,1	0,0440	0,0431	0,0422	0,0413	…	0,0371	0,0363
…	…	…	…	…	…	…	…
5,0	0,0000015	-	-	-	-	-	-

Расчет значений t может быть произведен с помощью стандартной функции Excel ***НОРМАЛИЗАЦИЯ.***

НОРМАЛИЗАЦИЯ(X';$\bar{X}$;σ)

НОРМАЛИЗАЦИЯ(95;124,20;13,69) → -2,133263057

Методы расчета значений теоретической нормированной функции $\varphi_{(t)}$

Microsoft Excel - Лекция_1_Пример_1.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч % 000

B11 = =КОРЕНЬ(C24)

	A	B	C	D
1	Середина интервала, x'	φ(t)		
2	95	0,040993401	=НОРМРАСП(A2;B10;B11;0)*B11	
3	105	0,149163377		
4	115	0,31828316		
5	125	0,398261492		
6	135	0,292230925		
7	145	0,125743988		
8	155	0,031728642		
9				
10	$\bar{x}$ =	124,2000000		
11	стандарт =	13,6879509		
12				

СП .

<u>ную функцию</u> <u>рас-</u>
<u>плотности распре-</u>

ной функции $\varphi_{(t)}$ не-
и **НОРМРАСП** на **σ**.

НОРМРАСП(95;124,20;13,69;0)* 13,69 → 0,041021332

Критерий согласия Пирсона

Критерий согласия Пирсона:

$$\chi^2 = \frac{\sum_{i=1}^{n} (f_i - f_{mi})^2}{\sum_{i=1}^{n} f_{mi}}$$

Для найденного значения критерия согласия Пирсона и числа степеней свободы γ=n-1 определяется соответствующая вероятность P(χ^2).

При P(χ^2)>0,5 считается, что эмпирическое и теоретическое распределения близки, при P(χ^2)$\in$[0,2;0,5] совпадение удовлетворительное, в остальных случаях – недостаточное.

Критерий согласия Пирсона

Способы нахождения вероятности $P(\chi^2)$.

1. С помощью таблиц распределения Пирсона (χ^2):

Число степеней свободы	Вероятность $P(\chi^2)$.						
	0,999	**0,995**	…	**0,70**	**0,50**	…	**0,001**
1	**0,05157**	**0,04393**	…	**0,148**	**0,455**	…	**10,827**
2	**0,00200**	**0,0100**	…	**0,713**	**1,386**	…	**13,815**
…	…	…	…	…	…	…	…
6	**0,381**	**0,676**	…	**3,828**	**5,348**	…	**22,457**
…	…	…	…	…	…	…	…
30	**11,588**	**13,787**	…	**25,508**	**29,336**	…	**59,703**

Для приведенного примера:

γ=7-1=6; χ^2 =4,37;

$P(\chi^2)$ находится в диапазоне от 0,5 до 0,7.

В линейном приближении $P(\chi^2)$=0,628.

Критерий согласия Пирсона

2. С помощью стандартной функции Excel ***ХИ2ТЕСТ***.

ХИ2ТЕСТ($\vec{f}$;$\vec{f}_m$).

Функция ХИ2ТЕСТ в качестве промежуточного действия вычисляет χ^2 и возвращает вероятность $P(\chi^2)$.

	A	B	C	D	E
1	**Число ткачих, f**	**fm**			
2	**2**	**3**			
3	**15**	**11**			
4	**20**	**23**			
5	**32**	**29**			
6	**18**	**21**			
7	**9**	**9**			
8	**4**	**2**	**P=**	**0,626779673**	**=ХИ2ТЕСТ(A2:A8;B2:B8)**
9					

Критерий согласия Пирсона

Рассчитав значение $P(\chi^2)$ можно получить значение критерия Пирсона с помощью стандартной функции Excel ***ХИ2ОБР***.

ХИ2ОБР($P(\chi^2)$; γ).

Функция ХИ2ОБР возвращает значение χ^2.

	A	B	C	D	E
1	**Число ткачих, f**	**fm**			
2	**2**	**3**			
3	**15**	**11**			
4	**20**	**23**			
5	**32**	**29**			
6	**18**	**21**			
7	**9**	**9**	**P=**	**0,626779673**	**=ХИ2ТЕСТ(A2:A8;B2:B8)**
8	**4**	**2**	**Критерий Пирсона=**	**4,369693646**	**=ХИ2ОБР(D7;6)**
9					

Критерий согласия Колмогорова

Критерий согласия Колмогорова:

$$\lambda = \frac{D}{\sqrt{\sum_{i=1}^{n} f_i}}$$

$$\lambda = \frac{3}{10} = 0{,}3$$

$P(\lambda) \approx 1$

где D – максимальное значение разности между накопленными эмпирическими и теоретическими частотами.

Эмпирические частоты f	Теоретические частоты f_m	Накопленные эмпирические частоты S_1	Накопленные теоретические частоты S_2	Отклонение $\lvert S_1-S_2 \rvert$
2	3	2	3	1
15	11	17	14	3
20	23	37	37	0
32	29	69	66	3
18	21	87	87	0
9	9	96	96	0
4	2	100	98	2

Критерий согласия Романовского

Критерий согласия Романовского:

$$C = \frac{\chi^2 - \gamma}{\sqrt{2\gamma}}$$

где χ^2 – критерий Пирсона, γ - число степеней свободы (γ=n-3).

При С<3 различие несущественно, что позволяет считать эмпирическое распределение близким к нормальному.