ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ

ФН-12. Магистры - 3 семестр

Семинар 4. АППРОКСИМАЦИЯ ТРИГОНОМЕТРИЧЕСКИМИ ПОЛИНОМАМИ. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В СИСТЕМЕ MATLAB.

Тригонометрический полином.

Пусть неизвестная функция ищется в виде тригонометрического полинома степени r с неизвестными коэффициентами α_j , β_j , рассматриваемого на $[0, T]$.
Тригонометрический полином степени r имеет вид

$$f(t, \alpha, \beta) = \alpha_0 + \sum_{j=1}^{r} (\alpha_j \cos(j\omega t) + \beta_j \sin(j\omega t)). \tag{4.1}$$

Модель в комплексной форме:

$$f(t, \Theta) = \sum_{j=-r}^{r} (\theta_j \phi_j(t)), \ \ \phi_j(t) = \exp(ij\omega t), \tag{4.2}$$

где $\theta_0 = \alpha_0$, $i = \sqrt{-1}$, $\theta_{\pm j} = \alpha_j \pm i\beta_j$, $j = \overline{1, r}$.
Если $\forall j, \ j = \overline{1, r}$, $\alpha_j \in \mathbb{R}$, $\beta_j \in \mathbb{R}$, то $\alpha_j = Re\theta_j$, $\beta_j = Im\theta_j$
Тригонометрическую регрессию применяют к периодическим функциям, заданным на сетке с равноотстоящими узлам: $t_l = (l-1)T/N$, (значение в узле $t_{n+1} = T$, в силу периодичности, совпадает со значением в начальном узле $t_{n+1} = 0$).

Реализация ДПФ и ОДПФ пакете MATLAB.

Используют FFT-алгоритм быстрого Фурье-преобразования.
fft (Fast Fourier Transform) Одномерное дискретное прямое преобразования Фурье:
Функция $Y = fft(X)$ вычисляет для вектора данных X ДПФ, если массив X двумерный, вычисляется ДПФ каждого столбца.
Функция $Y = fft(X, n)$ вычисляет n -точечное дискретное преобразование Фурье. Если $|X| < n$ – недостающие строки массива X заполняются нулями; если $|X| > n$ – лишние строки удаляются.
$ifft$ (Inverse Fast Fourier Transform) Одномерное дискретное обратное преобразования Фурье:
Функция $X = ifft(Y)$ вычисляет обратное преобразование Фурье для массива Y.
Функция $X = ifft(Y, n)$ вычисляет n-точечное обратное преобразование Фурье для массива Y.
Если $Z = iff(Y)$, то $b_j = Z_{j+1}$, $b_{-j} = Z_{N-j+1}$, $j = \overline{1, r}$.
При $r \geq N/2$ вектор Y восстанавливается точно ($Y = \hat{Y}$) (соответствует задаче тригонометрической интерполяции).

Задача 3.1. Построить вектор Y значений аппроксимирующего тригонометрического полинома степени r для функции $f(x)$, заданной на отрезке $[O,\, T]$, по измерениям в n точках со случайными ошибками $\varepsilon \to N(0, \sigma)$.

Алгоритм решения задачи.

1. Сформировать узлы сетки.
2. Получить вектор Y_0 – вектор значений функции $f(x)$.
3. Получить вектор Y – вектор "экспериментальных" значений с учетом аддитивной выходной помехи (зашумленный вектор Y_0) Y в узлах сетки.
4. Выполнить ДПФ "экспериментального" вектора Y. Полученный в результате ДПФ вектор содержит практически весь спектр частот.
5. Выполнить фильтрацию с целью убрать шумы. Выбрать из вектора Z элементы соответствующие коэффициентам тригонометрического полинома степени r, остальные компоненты вектора обнулить. Сформировать вектор $Z1$ состоящих из элементы вектора Z и 0 можно, например, используя данный оператор $Z1 = Z. * ((1 : n <= r + 1)|(1 : n >= n - r + 1))$.
6. Выполнить ОДПФ и получить значения тригонометрического полинома.
7. Результаты моделирования отобразить на графиках.
8. Оценить уровень шумов (вычислить СКО).
9. Оценить ошибку аппроксимации (вычислить СКО восстановленного сигнала от вектора значений функции $f(x)$).

Исходные данные: $f(t) = 1/(1 + (t - T/2)4)$,
$r = 3$, $T = 2$, $n = 100$, $\sigma = 0.05$

Значения коэффициентов тригонометрического полинома:
$\theta_0 = 85.4242$,
$\theta_1 = (-9.6371 - 0.0775i)$, $\theta_{-1} = -9.6371 + 0.0775i$
$\theta_2 = 4.2641 + 0.3440i$, $\theta_{-2} = 4.2641 - 0.3440i$
$\theta_3 = -1.1802 - 0.0790i$, $\theta_{-3} = -1.1802 + 0.0790i$

Оценка уровня шумов $\sigma = 0.0462$
Ошибка аппроксимации $\Delta = 0.0223$

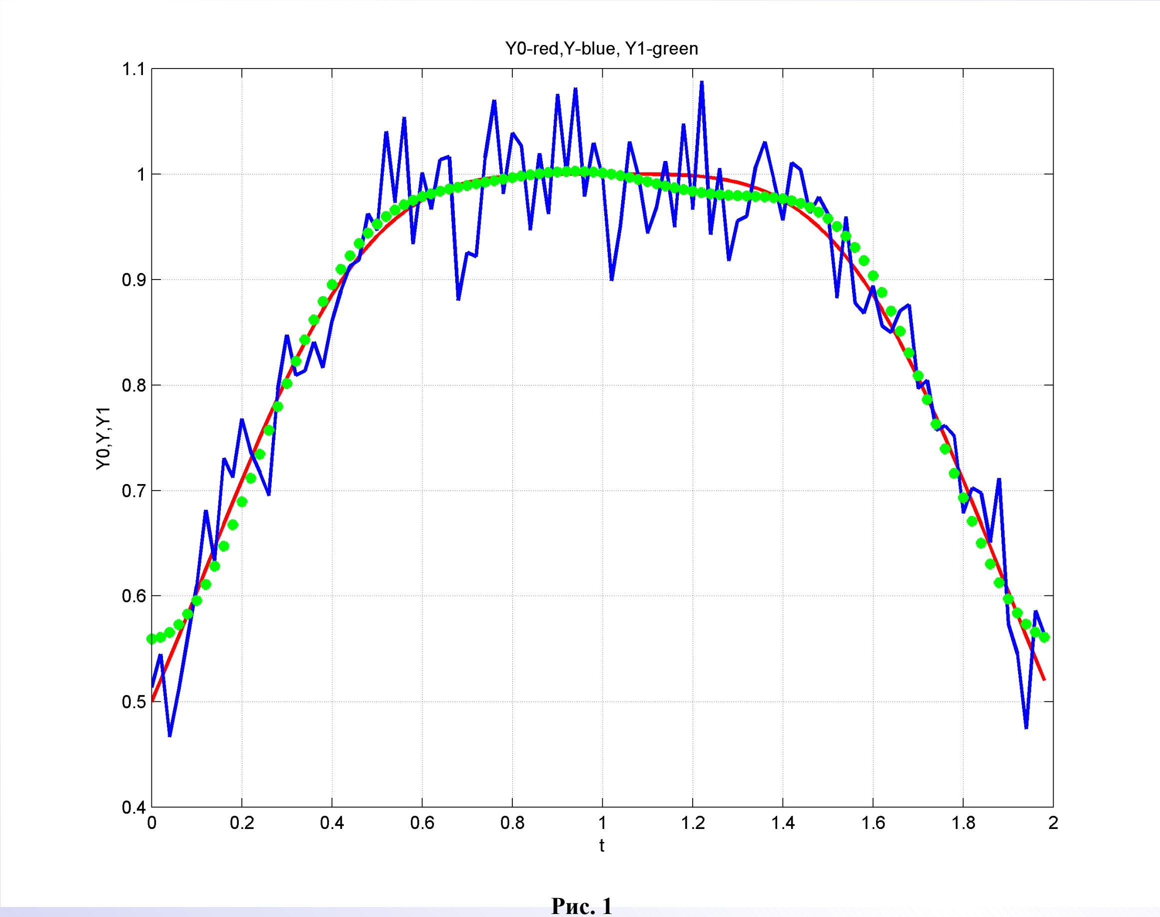
Y0-red,Y-blue, Y1-green
Y0,Y,Y1
t
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Рис. 1

$$r = 9$$

Оценка уровня шумов $\sigma = 0.0537$
Ошибка аппроксимации $\Delta = 0.0224$

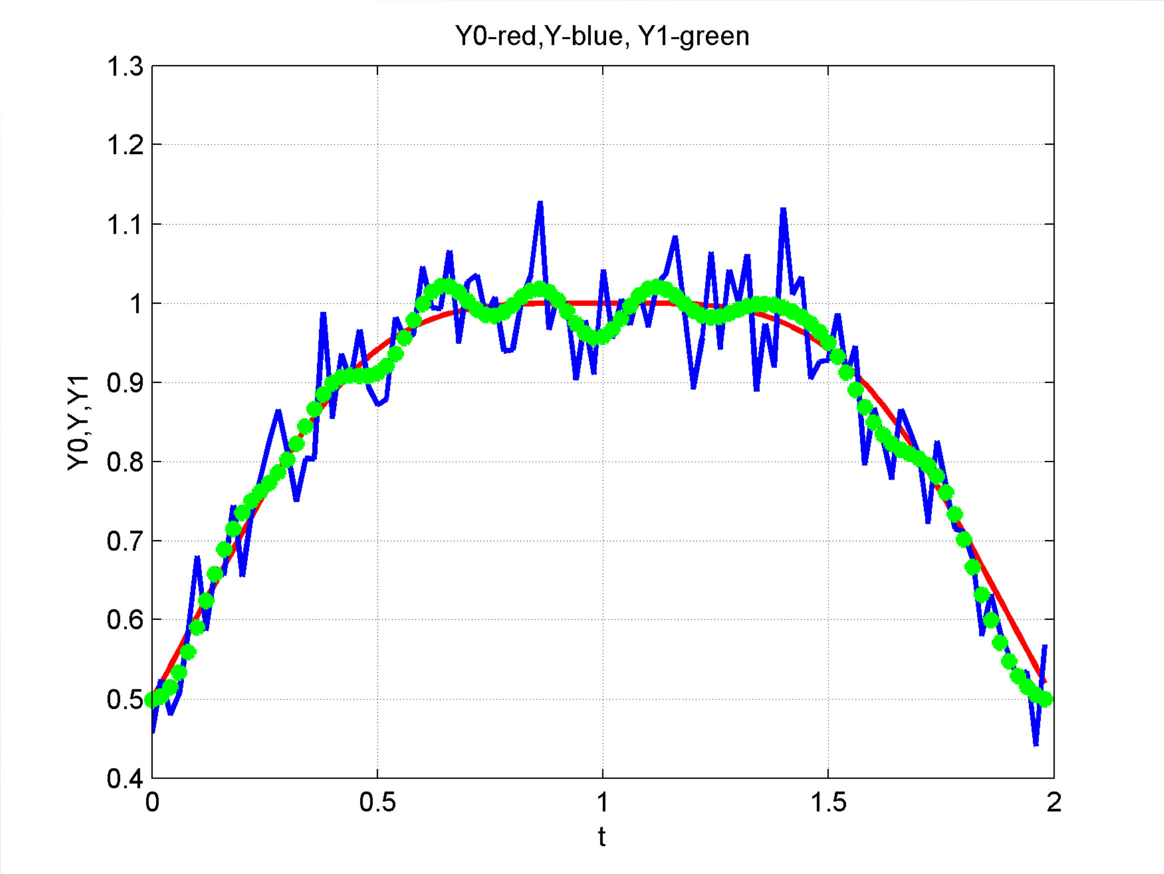

Рис. 2

Задание

1. Задать новую функцию $f(t)$, не являющуюся тригонометрическим или стандартным полиномом.
2. Задать отрезок $[0, T]$, на котором производится аппроксимация.
3. Положить количество измерений $n = 100$.
4. Вычислить узлы сетки, соответствующие заданной величине n.
5. Вычислить значения функции $f(t)$ в узлах построенной сетки: $y_k = f(t_k)$.
6. Задать значения СКО для моделирования ошибок измерений $\varepsilon \to N(0, \sigma)$ /
7. Получить сеточную функцию $z_k = y_k + \varepsilon_k$, моделирующую ”зашумленные” измерения.
8. Задать степень r тригонометрического полинома и получить аппроксимацию функции $f(t)$.
9. Исследовать влияние r на оценка ошибки аппроксимации.
10. Выбрать оптимальные значения r.
11. Для данных полученных в п.7 получить полиномиальную аппроксимации. Обосновать выбор степени r.

Статистическое моделирование в системе MATLAB.

Проверка гипотезы о нормальном распределении генеральной совокупности

Критерий согласия для проверки гипотезы о законе распределения исследуемой случайной величины. Критерий проверки гипотезы о предполагаемом законе неизвестного распределения называют критерием согласия.

Позволяет оценить значимость различий между фактическим количеством исходов (качественных характеристик выборки), и теоретическим количеством, которое можно ожидать при справедливости нулевой гипотезы/

Примеры критериев согласия: χ^2 (хи квадрат) К. Пирсона, Колмогорова, Смирнова.

Критерий согласия Пирсона (критерий согласия χ^2) наиболее часто употребляемый для проверки гипотезы о принадлежности выборки теоретическому закону распределения.

В качестве критерия проверки основной гипотезы используется случайную величину: $\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i^0)^2}{n_i}$, где n_i — эмпирические частоты, полученные из выборки, n_i^0 — теоретические частоты, s — число групп (интервалов) .

Алгоритм критерия согласия Пирсона.

Выбор теоретического закона распределения.

Оценить параметры распределения по выборке.

Вычислить теоретические значения частот (через теор. вероятности попадания в интервал) и сравнить с эмпирическими (наблюдаемыми).

Рассчитать значение статистики $\chi^2_{набл}$, сравнить с критическим значением $\chi^2_{крит}$ со степенями свободы $k = s - 1 - r$, s — число групп выборки, r — число параметров предполагаемого закона распределения.

$\chi^2_{набл} < \chi^2_{крит}$ — нулевая гипотеза о законе распределения принимается.

Для применения критерия согласия Пирсона (χ^2) требуется выборка размера $n > 50$.

Дано: экспериментальные данные по оценке анеуплоидии в культурах мультипотентных мезенхимных стволоваых клеток (МСК) костного мозга.
Задача. Проверить гипотезу о нормальном распределении генеральной совокупности.
Обозначения:
n_j всего клеток в культуре j ($j = 1, \ldots, 17$); v_0 количество нульхромосомных клеток; v_1 количество монохромосомных клеток; v_2 количество дихромосомных клеток (нормальных); v_3 количество трихромосомных клеток; v_4 количество тетрахромосомных клеток

Проверка гипотезы о нормальном распределении ГС с использованием критерия асимметрии S_s и эксцесса E_s.

Третий центральный момент μ_3 характеризует степень асимметрии кривой распределения относительно математического ожидания.

Безразмерная величина $A_s = \frac{\mu_3}{\sigma^3}$, называется коэффициентом асимметрии теоретического распределения.

Безразмерная величина $S_s = \frac{\mu_3}{S^3}$, называется коэффициентом асимметрии эмпирического распределения.

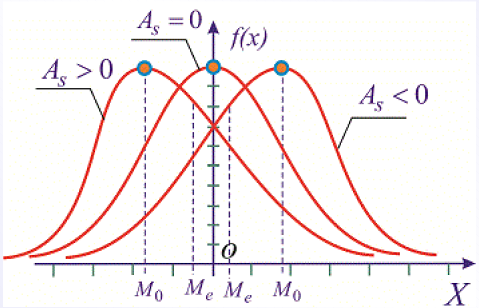

Рис. 3

Асимметрия нормального распределения равна 0. Чем меньше по модулю A_s, тем рассматриваемое эмпирическое распределение ближе к нормальному распределению. $A_s > 0$ —распределение "скошено" вправо, $A_s < 0$ то влево.

Четвертый центральный момент μ_4 определяет эксцесс (островершинность кривой распределения). За характеристику этого свойства принимают безразмерную величину г,

Безразмерная величина $\varepsilon_s = \frac{\mu_4}{\sigma^4} - 3$, называется коэффициентом эксцесса теоретического распределения.

Безразмерная величина $E_s = \frac{\mu_3}{S^3}$, называется коэффициентом эксцесса эмпирического распределения.

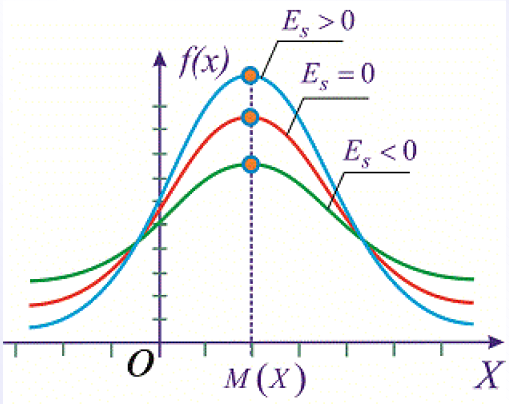

Рис. 4

Эксцесс нормального распределения равен 0. $E_s > 0$ $(E_s < 0)$ — эмпирическое распределение более островершинное (более низкое и пологое) (по сравнению с нормальным распределением).

Оценка существенности асимметрии A_s и эксцесса E_s проводится с помощью среднеквадратической ошибки: $\sigma A_s = \sqrt{\dfrac{6(n-1)}{(n+1)(n+3)}}$,

$$\sigma E_s = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}},$$

где n — число наблюдений.

$\dfrac{|A_s|}{\sigma A_s} < 3$ — асимметрия не существенна, ее наличие обусловлено влиянием случайных факторов.

$\dfrac{|A_s|}{\sigma A_s} \geq 3$ асимметрия существенна.

$\dfrac{|E_s|}{\sigma E_s} < 3$ — эксцесс не существен, его наличие обусловлено влиянием случайных факторов.

$\dfrac{|E_s|}{\sigma E_s} \geq 3$ — эксцесс существен.

Проверка гипотезы о нормальном распределении ГС с использованием критерия асимметрии S_k и эксцесса E_k.

Абсолютные значения							Относительные значения (частоты)				
№(j)	n_j	v_0	v_1	v_2	v_3	v_4	v_0/n_j	v_1/n_j	v_2/n_j	v_3/n_j	v_4/n_j
1	1368	0	7	1327	8	26	0	0,0051	0,9700	0,0058	0,01901
2	1019	0	6	1001	4	8	0	0,0059	0,9823	0,00393	0,00785
3	1039	0	13	1016	3	7	0	0,01251	0,9779	0,00289	0,00674
4	323	0	3	318	0	2	0	0,00929	0,9845	0	0,00619
5	1034	0	17	1007	5	5	0	0,01644	0,9739	0,00484	0,00484
6	1012	0	5	1005	1	1	0	0,00494	0,9931	0,00099	0,00099
7	1013	0	6	1005	1	1	0	0,00592	0,9921	0,00099	0,00099
8	901	0	4	890	5	2	0	0,00444	0,9878	0,0056	0,0022
9	1024	0	7	1013	2	2	0	0,00664	0,9893	0,0019	0,0019
10	1051	0	10	1014	2	25	0	0,00951	0,0019	0,0019	0,02379
11	411	0	4	397	6	4	0	0,00973	0,9659	0,0146	0,0973
12	1024	0	1	1023	0	0	0	0,00098	0,9990	0	0
13	934	0	0	931	3	0	0	0	0,9968	0,0032	0
14	583	0	3	569	0	1	0	0,00524	0,9930	0	0,00175
15	668	0	3	662	3	0	0	0,00449	0,991	0,0045	0
16	1081	3	32	1042	3	1	0,0028	0,0296	0,964	0,0028	0,00093
17	1025	0	10	1015	0	0	0	0,0098	0,9902	0	0

№(j)	n_j	v_0	v_1	v_2	v_3	v_4
MAX	1368	3	32	1327	8	26
MIN	323	0	0	318	0	0
MAX-MIN	1045	3	32	1009	8	26
SUMM	1550	3	131	15235	46	85
AVERAGE	911,8	0,176	7,71	896,2	2,706	5

частоты	v_0/n_j	v_1/n_j	v_2/n_j	v_3/n_j	v_4/n_j
MAX	0,0028	0,0296	0,9990	0,0146	0,0238
MIN	0	0	0,9639	0	0
MAX-MIN	0,0028	0,0296	0,0351	0,014599	0,0238
SUMM	0,0028	0,1407	16,716	0,0540	0,0869
AVERAGE	0,0002	0,0083	0,983	0,0032	0,0051
D[X]	4,53E-07	4,63E-05	0,00014	1,26E-05	4,75E-05
σ	0,000673	0,00680	0,01166	0,003549	0,00689
S_k (асимметрия)	3,638	1,8125	-0,47	1,926	1,616
E_k (эксцесс)	14,176	6,413	1,734	6,925	4,518

Оценка ассиметрии $\sigma A_s = \sqrt{\dfrac{6(17-1)}{(17+1)(17+3)}} = 0,516$.

Оценка эксцесса $\sigma E_s = \sqrt{\dfrac{24 \cdot 17(17-2)(17-3)}{(17-1)^2(17+3)(17+5)}} = 0,872$.

Рассмотрим v_0/n_j :

$\dfrac{3,638}{0,516} = 7,05 \geq 3$ асимметрия существенна.

$\dfrac{14,176}{0,872} = 16,257 \geq 3$ — эксцесс существенен.

Гипотеза о нормальном распределении данных отвергается, т.к. асимметрия и эксцесс не равны их оценкам.

Проверка на принадлежность одной генеральной совокупности экспериментальных данных.

Проверка на однородность (принадлежность одной генеральной совокупности(ГС)) двух выборок.

Для увеличения количества экспериментальных данных предлагается объединить две выборки $x_1,\ x_2 \ldots x_{n_1}$ и $y_1,\ y_2 \ldots y_{n_2}$, полученные в результате двух экспериментов.

Объединение экспериментальных данных возможно в том случае, если обе выборки получены из одной генеральной совокупности.

Критерий χ^2

Используется для проверки однородности данных, имеющих дискретную структуру, т.е. когда в опытах наблюдается некоторый переменный признак, принимающий конечное число (s) различных значений.

Осуществлено k последовательных серий независимых испытаний, состоящих из n_1 , .. , n_j наблюдений соответственно.

В каждом опыте наблюдается некоторый переменный признак, принимающий одно из s значений (исходов).

Обозначим:

$n_1, .., n_j$ количество наблюдений в каждой серии ($j = 1..k$);

n — общее число наблюдений $n = \sum n_j$;

v_{ij} – число реализаций i - го исхода в j ой серии $\sum v_{ij} = n_j$, ($i = 1 \ldots s, j = 1 \ldots k$);

p_{ij} вероятность появления i -го исхода в испытаниях j й серии, ($i = 1 \ldots s, j = 1 \ldots k$);

p_i вероятность появления i -го исхода, одинаковая для любой серии испытаний;

$p = (p_1, \ldots, p_s)$ некоторый неизвестный вектор вероятностей, $(p_1 + \ldots + p_s) = 1$

Требуется проверить гипотезу H_0 о том, что все наблюдения производились над одной и той же случайной величиной.

$H_0 : (p_{1j}, \ldots, p_{sj}) = (p_1, \ldots, p_s)$.

Математическое ожидание величины v_{ij} при выполнении гипотезы H_0 $M(\nu_{ij}|H_0) = n_j p_i$. В качестве меры отклонения опытных данных от средних выберем статистику $\chi^2_n(p) = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\nu_{ij} - n_j p_i)^2}{n_j p_i}$.

Неизвестные параметры $(p_1, \ldots, p_s)$ можно методом максимального правдоподобия $L(p) = c\prod p_j^{v_i}$, где $v_i = \sum_{j=1}^{k} v_{ij}$, и ее логарифм. Оценки $\hat{p}_i = \frac{\nu_i}{n}$.

$$\chi^2_n(p) = n\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\nu_{ij} - n_j \nu_i)^2}{n_j \nu_i} .$$

Критическую область зададим в виде $\mathfrak{F}_{1\alpha} = \{t \geq t_\alpha\}$, $t_\alpha = \chi^2((s-1),(k-1))$

Проверяется выборка, состоящая из k разных культур клеток, в каждой культуре подсчитывается
— общее число клеток (n_j),
— количество нульхромосомных клеток (исход v_0);
— количество монохромосомных клеток (исход v_1);
— количество дихромосомных клеток (исход v_2);
— количество трихромосомных клеток (исход v_3);
— количество тетрахромосомных клеток (исход v_4).
$k = 17$, $s = 5$.
Опыт здесь понимается как подсчет клеток определенного типа(v_i) в j ой культуре ($i = 1 \dots s$, $j = 1 \dots k$).

№(j)	n_j	v_0	v_1	v_2	v_3	v_4	ИТОГ
1	1368	0	7	1327	8	26	
2	1019	0	6	1001	4	8	
3	1039	0	13	1016	3	7	
4	323	0	3	318	0	2	
5	1034	0	17	1007	5	5	
6	1012	0	5	1005	1	1	$\{\alpha = 0.05\},$
7	1013	0	6	1005	1	1	
8	901	0	4	890	5	2	$\chi^2_\alpha((s-1)(k-1)) =$
9	1024	0	7	1013	2	2	$= \chi^2_\alpha(4 \cdot 16) =$
10	1051	0	10	1014	2	25	$= \chi^2_\alpha(64) = 83.675$
11	411	0	4	397	6	4	
12	1024	0	1	1023	0	0	$\chi^2_{набл} = 520.7$
13	934	0	0	931	3	0	$520,70 \geq 83,675$
14	583	0	3	569	0	1	
15	668	0	3	662	3	0	
16	1081	3	32	1042	3	1	
17	1025	0	10	1015	0	0	

Гипотезу однородности H_0 отвергаем, наблюденное значение $\chi^2_{набл}$ удовлетворяют неравенству $\chi^2_{набл} > \chi^2_\alpha$. Вероятность ошибочно отклонить истинную гипотезу приблизительно равна 5% ($\alpha = 0,05$).

Критерий Вилкоксона

1945 г.: Вилкоксон опубликовал критерий сравнения выборок одинакового объема, 1947 г.: Манн и Уитни обобщили критерий на выборки различного объема.

Непараметрический статистический критерий. Используется для проверки однородности двух независимых выборок $x_1, \; x_2 \ldots x_{n_1}$ и $y_1, \; y_2 \ldots y_{n_2}$. Применим к случайным величинам, распределение которых неизвестно.

Если выборки однородны, считают, они извлечены из одной и той же ГС, имеют одинаковые (неизвестные), непрерывные функции распределения $F_1(x)$ и $F_2(x)$.

Нулевая гипотеза $H_0: \; F_1(x) = F_2(x)$ — две выборки однородны, при всех значениях аргумента функции распределения равны между собой.

Конкурирующие гипотезы: $F_1(x) \neq F_2(x)$, $F_1(x) < F_2(x)$, $F_1(x) > F_2(x)$.

Гипотеза $F_1(x) < F_2(x)$ означает $X > Y$

($(F_1(x) < F_2(x)) \equiv (P(X < x) < P(Y < x)) \Rightarrow$

$\Rightarrow (P(X > x) > P(Y > x)) \Rightarrow (X > Y)$).

$X > Y$: вероятность того, что $X > x$, больше, чем вероятность $Y > x$.

Непараметрические методы применяются, когда объем выборок мал (например, $n < 50$). При $n > 50$ смысла использовать непараметрические статистики нет.

Алгоритм

I. $n_1 \leq 25 \& n_2 \leq 25$, $n_1 \leq n_2$.

1. H_0 : $F_1(x) = F_2(x)$, $F_1(x) \neq F_2(x)$.

1) расположить варианты обеих выборок в возрастающем порядке, в виде одного вариационного ряда и найти в этом ряду наблюдаемое значение критерия $W_{набл}$ — сумму порядковых номеров вариант **первой** выборки ($n_1 \leq n_2$);

2) найти по табл. Критические точки критерия Вилкоксона нижнюю критическую точку $w_{нижн.кр}(Q; n1, n2)$, где $Q = \alpha/2$;

3) найти верхнюю критическую точку по формуле $w_{верх.кр.} = (n_1 + n_2 + 1)n_1 - w_{нижн.кр}$.

$W_{нижн.кр.} < W_{набл.} < W_{верхн.кр.}$ — нет оснований отвергнуть нулевую гипотезу.

$W_{набл.} < W_{нижн.Кр.}$ или $W_{набл.} > W_{верхн.кр.}$ — нулевую гипотезу отвергают.

2. $H_0: \ F_1(x) = F_2(x)$, $F_1(x) > F_2(x)$.
1) расположить варианты обеих выборок в возрастающем порядке, в виде одного вариационного ряда и найти в этом ряду наблюдаемое значение критерия $W_{набл}$ — сумму порядковых номеров вариант **первой** выборки ($n_1 \leq n_2$);
2) Найти по таблице нижнюю критическую точку $w_{нижн.кр}(Q; n1, n2)$, где $Q = \alpha$;
$W_{нижн.кр.} < W_{набл.}$ — нет оснований отвергнуть нулевую гипотезу.
$W_{набл.} < W_{нижн.Кр.}$ — нулевую гипотезу отвергают.

3. $H_0: \ F_1(x) = F_2(x)$, $F_1(x) < F_2(x)$.
1) расположить варианты обеих выборок в возрастающем порядке, в виде одного вариационного ряда и найти в этом ряду наблюдаемое значение критерия $W_{набл}$ — сумму порядковых номеров вариант **первой** выборки ($n_1 \leq n_2$);
2) Найти по таблице верхнюю критическую точку $w_{верх.кр.}(Q; n1, n2) = (n_1 + n_2 + 1)n_1 - w_{нижн.кр}(Q; n1, n2)$, где $Q = \alpha$;
$W_{набл.} < W_{верхн.кр.}$ — нет оснований отвергнуть нулевую гипотезу.
$W_{набл.} > W_{верхн.кр.}$ — нулевую гипотезу отвергают.
Замечание. Если несколько вариант только **одной** выборки совпадают, то в общем вариационном ряду совпавшие варианты нумеруют так, как если бы они были различными числами (приписывают обычные порядковые номера); Если же совпадают варианты **разных** выборок, им всем присваивают один и тот же порядковый номер, равный среднему арифметическому порядковых номеров, которые имели бы эти варианты до совпадения.

II. $(n_1 \le 25) OR (n_2 \le 25)$, $n_1 \le n_2$.

1. $H_0: \ F_1(x) = F_2(x)$, $F_1(x) \ne F_2(x)$.

1) расположить варианты обеих выборок в возрастающем порядке, в виде одного вариационного ряда и найти в этом ряду наблюдаемое значение критерия $W_{набл}$ — сумму порядковых номеров вариант **первой** выборки ($n_1 \le n_2$);

2) Нижняя критическая точка

$$W_{нижн.кр.}(Q; n1, n2) = \left[\frac{(n_1 + n_2 + 1)n_2 - 1}{2} - z_{кр}\sqrt{\frac{(n_1 + n_2 + 1)n_1 n_2}{12}} \right]. \tag{4.3}$$

Здесь $Q = \alpha/2$, знак $[a]$ — целая часть a, $z_{кр}$ находят по табл. функции Лапласа по равенству $\Phi(z_{кр}) = (1 - \alpha)/2$.

3) найти верхнюю критическую точку по формуле $w_{верх.кр.} = (n_1 + n_2 + 1)n_1 - w_{нижн.кр}$.

$W_{нижн.кр.} < W_{набл.} < W_{верхн.кр.}$ — нет оснований отвергнуть нулевую гипотезу.

$W_{набл.} < W_{нижн.Кр.}$ или $W_{набл.} > W_{верхн.кр.}$ — нулевую гипотезу отвергают.

2. $H_0: \quad F_1(x) = F_2(x)\,, F_1(x) > F_2(x)\,, \quad F_1(x) < F_2(x)\,.$ нижнюю критическую точку находят по формуле (4.3), $Q = \alpha\,,$ $z_{\kappa p}$ находят по таблице функции Лапласа по равенству $\Phi(z_{\kappa p}) = (1-2\alpha)/2\,,$ пп. 1,3, сохраняются.

ПРИМЕР.

Можно ли объединить две выборки, полученные в результате двух экспериментов ?

Исходные данные. Имеются две выборки. Первая выборка $\{Y\}$ содержет количество дихромосомных клеток на ранних пассажах (17 элементов), вторая $\{X\}$ — на поздних пассажах (11 элементов).

Проверить с уровнем значимости $\alpha = 0,05$, что выборки однородны, т.е. они извлечены из одной и той же ГС и имеют одинаковые (неизвестные), функции распределения $(F_1(x)$ и $F_2(x))$

	1	2	3	4	5	6	7	8	9
Y	0,9700	0,9823	0,9779	0,9845	0,9739	0,9931	0,9921	0,9878	0,9893
X	0,965	0,9882	0,9853	0,9921	0,9841	0,9777	0,9990	0,9806	0,9883

	10	11	12	13	14	15	16	17
Y	0,9648	0,9659	0,9990	0,9968	0,9930	0,9910	0,9639	0,9902
X	0,9892	0,9577	—	—	—	—	—	-

$H_0: \; F_1(x) = F_2(x)$, $F_1(x) \neq F_2(x)$, $\alpha = 0,05$.

По условию критерия: $n_1 < n_2$.

Обозначения: выборка $\{X\}$ – выборка № 1 выборка $\{Y\}$ – выборка № 2.

$n_1 = 11$, $n_2 = 17$, $n_1 < n_2$. $n_1 < 25$, $n_2 < 25$.

Расположем исходные данные в таблицу в соответствии с условиями критерия Вилкоксона.

№	ν_2/n_j	№(mix)	ν_2/n_j	RANGE	$X\ (I)$	$Y\ (II)$	Σ_X	Σ_Y
I 16	0,9639	II 11	0,9577	1	x		1	2
I 10	0,9648	I 16	0,9639	2		Y	4	3
I 11	0,9659	I 10	0,9648	3		Y	8	5
I 1	0,9700	II 1	0,9650	4	x		10	6
I 5	0,9739	I 11	0,9659	5		Y	12	7
I 3	0,9779	I 1	0,9700	6		Y	14	9
I 2	0,9823	I 5	0,9739	7		Y	16	11
I 4	0,9845	II 6	0,9777	8	x		17	13
I 8	0,9878	I 3	0,9779	9		Y	18	15
I 9	0,9893	II 8	0,9806	10	x		22	19
I 17	0,9902	I 2	0,9823	11		Y	27	20
I 15	0,9910	II 5	0,9841	12	x		—	21
I 7	0,9921	I 4	0,9845	13		Y	$W_x = 149$	23
I 14	0,9930	II 3	0,9853	14	x		149	24

$W_{набл}$ — Сумма порядковых номеров вариант **первой** выборки ($n_1 \leq n_2$)

Исходные данные (продолжение).

№	ν_2/n_j	№(mix)	ν_2/n_j	RANGE	$X\ (I)$	$Y\ (II)$	Σ_X	Σ_Y
I 6	0,9931	I 8	0,9878	15		Y	—	25
I 13	0,9968	II 2	0,9882	16	X		—	26
I 12	0,9990	II 9	0,9883	17	X		—	28
II 11	0,9577	II 10	0,9892	18	X			
II 1	0,965	I 9	0,9893	19		Y	—	—
II 6	0,9777	I 17	0,9902	20		Y	$W_x = 149$	$W_y = 257$
II 8	0,9806	I 15	0,9910	21		Y	—	
II 5	0,9841	II 4	0,9921	22	X		149	257
II 3	0,9853	I 7	0,9921	23		Y	$n_1 = 11$	$n_2 = 17$
II 2	0,9882	I 14	0,9930	24		Y		
II 9	0,9883	I 6	0,9931	25		Y		
II 10	0,9892	I 13	0,9968	26		Y		
II 4	0,9921	II 7	0,9990	27	X			
II 7	0,9990	I 12	0,9990	28		Y		

$W_{набл.} = W_x = 149$ б $n_1 = 11$, $n_2 = 17$.
$W_{нижн.кр.} = W(\alpha/2, n_1, n_2) = W(0,025, 11, 17) = 117$
$w_{верх.кр.}(Q, n1, n2) = (n_1 + n_2 + 1)n_1 - w_{нижн.кр} = (11 + 17 + 1)11 - 117 = 202$, где $Q = \alpha = 0,05$;
Условие $W_{нижн.кр.} < W_{набл.} < W_{верхн.кр.}$ выполнено ($117 < 149 < 202$), *нет оснований отвергнуть нулевую гипотезу* об однородности выборок

Задача 3.2.
Используя генератор случайных чисел, получить две выборки распределенные по нормальному закону, проверить выборки получены ли выборки из одной ГС или нет, математические ожидания и дисперсии взять равными. Повторить эксперимент для случая, когда выборки имеют разные математические ожидания и дисперсии.