

Институт Транспорта и Связи  
Заочное отделение

ЛАБОРАТОРНАЯ РАБОТА №3

по дисциплине

“МЕТОДЫ КОМПЬЮТЕРНОЙ ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ”

Тема: “Проверка однородности двух выборок”

Вариант №14

Выполнил: ст. Козлов С. А.

ст. код. 34524

24 июня 2002 г.

Проверил: доцент Люмкис В. Д.

Рига 2002

1. Согласно номеру варианта набрать данные двух выборок. Учесть, что в зависимости от вида критерия проверки однородности, данные обеих выборок должны быть расположены в одной переменной подряд, при этом вторая переменная должно содержать номер выборки, которой принадлежит данное значение. При другом варианте расположения данных выборки должны быть расположены в разных переменных.

Заданные совокупности помещены в одну переменную VAR1, а в переменную NEWVAR помещены группировочные коды 1 и 2, соответственно:

	VAR1	NEWVAR
1	11,110	1,000
2	8,925	1,000
3	11,775	1,000
4	11,664	1,000
5	8,571	1,000
6	9,494	1,000
7	10,749	1,000
8	10,839	1,000
9	9,658	1,000
10	11,006	1,000
11	10,732	1,000
12	10,735	1,000
13	9,499	1,000
14	9,852	1,000
15	9,481	1,000
16	9,519	1,000
17	10,102	1,000
18	8,505	1,000
19	10,145	1,000
20	8,583	1,000

	VAR1	NEWVAR
21	10,040	1,000
22	10,243	1,000
23	10,424	1,000
24	9,567	1,000
25	8,646	1,000
26	9,995	1,000
27	10,469	1,000
28	8,916	1,000
29	10,545	1,000
30	9,575	1,000
31	11,436	2,000
32	9,028	2,000
33	11,793	2,000
34	9,239	2,000
35	11,009	2,000
36	10,728	2,000
37	9,154	2,000
38	7,725	2,000
39	10,865	2,000
40	10,124	2,000

	VAR1	NEWVAR
41	11,132	2,000
42	10,978	2,000
43	10,855	2,000
44	10,584	2,000
45	11,525	2,000
46	11,882	2,000
47	10,509	2,000
48	9,324	2,000
49	8,515	2,000
50	10,402	2,000
51	11,112	2,000
52	11,583	2,000
53	11,863	2,000
54	10,254	2,000
55	9,631	2,000
56	11,210	2,000
57	9,441	2,000
58	10,279	2,000
59	8,560	2,000
60	11,205	2,000

2. Проверить выборки на однородность с помощью критерия Стьюдента.

Формула критерия Стьюдента (t-критерия) следующая:

$$t(n_1 + n_2 - 2) = \frac{\bar{x}_1(n_1) - \bar{x}_2(n_2)}{\tilde{s} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}};$$

где  $\bar{x}_1(n_1)$  и  $\bar{x}_2(n_2)$  - выборочные средние первой и второй выборки,  $\tilde{s}^2$  - оценка дисперсии, составленная из оценок дисперсии для каждой группы данных:

$$\mathfrak{F}^2 = \frac{1}{n_1+n_2-2} [(n_1-1)\bar{s}_1^2(n_1) + (n_2-1)\bar{s}_2^2(n_2)];$$

$$\tilde{s}_j^2(n) = \frac{1}{n_j-1} \sum_{i=1}^n (x_i - x_j(n))^2, \forall j = 1, 2.$$

Далее представлен результат обработки заданных совокупностей пакетом Statistica:

```
STAT.      Grouping: NEWVAR (laba3-2.sta)
BASIC      Group 1: G_1:1
STATS      Group 2: G_2:2
```

Variable	Mean G_1:1	Mean G_2:2	t-value	df	p
VAR1	9,978800	10,39817	-1,61661	58	,111389

По Probability Calculator  $t(58)=2,001717$   $p=,050000$ . Соответственно область принятия решения:  $(-2,001717, +2,001717)$ . Значение критерия попадает в ОПГ, следовательно, принимаем основную гипотезу: об однородности выборок.

3. Проверить выборки на однородность с помощью критерия Манна-Уитни (Вилкоксона).

Критерий Манна-Уитни (Вилкоксона) основан на статистике:

$$Z = \frac{U - \frac{1}{2}n_1n_2}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}},$$

$$\text{где: } U = R_1 - \frac{n_1(n_1+1)}{2},$$

а  $R_1$  – сумма всех рангов в объединённом вариационном ряде.

При  $n_1, n_2 > 10$  закон распределения статистики  $Z$  близок к  $N(0, 1)$ .

Далее представлен результат обработки заданных совокупностей пакетом Statistica:

```
STAT.      Mann-Whitney U Test (laba3-2.sta)
NONPAR     By variable NEWVAR
STATS      Group 1: 1 Group 2: 2
```

variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-level
VAR1	791,0000	1039,000	326,0000	-1,83327	,066772

По Probability Calculator  $Z(0;1)=1,959964$   $p=,0500000$ , а область принятия решения  $(-1,959964, +1,959964)$ . Полученное значение критерия попадает в ОПГ, следовательно, принимаем основную гипотезу: об однородности выборок.

4. Проверить выборки на однородность с помощью критерия Колмогорова-Смирнова.

Критерий Колмогорова-Смирнова основан на статистике:

$$D_{n_1, n_2} = \max_{-\infty < x < +\infty} |F_{n_1}(x) - F_{n_2}(x)|,$$

$F_n(x)$  - эмпирические функции распределения, построенные по данным выборок после их упорядочения (вариационные ряды).

Область принятия гипотезы  $(0, K_{1-\alpha} \cdot \sqrt{\frac{1}{n_0}})$ ,  $n_0 = \frac{n_1 n_2}{n_1 + n_2}$ ,  $K_{1-\alpha} \approx \sqrt{-\frac{\ln \alpha}{2}}$ .

Согласно задания:

$\alpha = 0,05$ ;

$K_{0,95} = 1,224$ ;

$n_0 = 15$

область принятия гипотезы:  $(0, 0,316)$ .

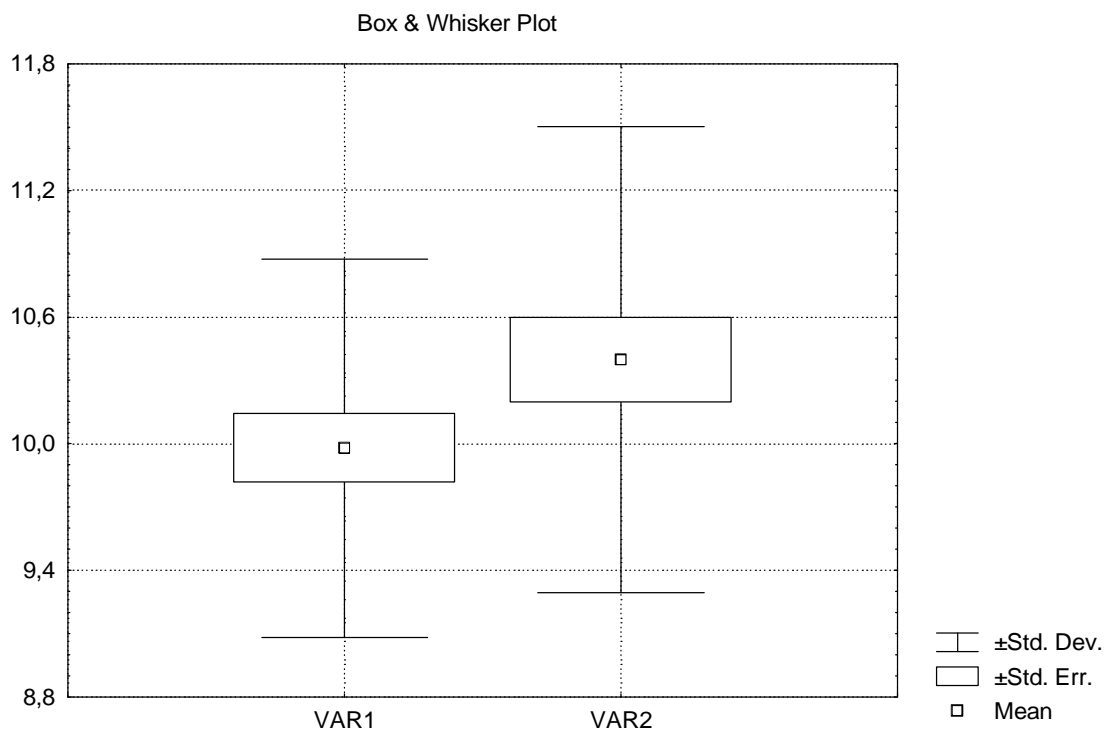
Далее представлен результат обработки заданных совокупностей пакетом Statistica:

```
STAT.      Kolmogorov-Smirnov Test (laba3-2.sta)
NONPAR     By variable NEWVAR
STATS      Group 1: 1 Group 2: 2
```

	Max Neg	Max Pos		Mean	Mean	Std.Dev.
variable	Differnc	Differnc	p-level	Group 1	Group 2	Group 1
VAR1	-,333333	,066667	p < .10	9,978800	10,39817	,895269

Полученное значение критерия  $-0.333333$  попадает в критическую область, следовательно, принимаем альтернативную гипотезу: о неоднородности выборок. ■

5. Исследовать различие в распределении выборок на графиках типа Box-Whisker.



Несмотря на сильное перекрытия стандартных отклонений, на приведённой диаграмме, из-за близкого расположения ошибок среднего можно говорить о спорном случае, или о граничном случае, в котором визуальная оценка невозможна и требуется оценка с помощью численных критериев.

О некоторой “неудачности” выборок говорит и тот факт что значения критериев располагались близко к квантилям, со стороны ОНР, как в случаях проверок по критериям Стьюдента и Манна-Уитни, или со стороны КО, при проверке по критерию Колмогорова-Смирнова.

С учётом того что приближения принятые в критерии Колмогорова-Смирнова справедливы при выборках с  $n > 50$ , необходимо руководствоваться табличными значениями квантилей при  $n = 15$ , не доступные автору работы, а так же приняв во внимание результаты проверок по критериям Стьюдента и Манна-Уитни можно сделать вывод об однородности выборок. На практике, в случае критических решений, можно было бы порекомендовать провести дополнительные исследования с большим количеством данных.

24 июня 2002 г.

/Козлов С. А./