Проверочная работа по статистике

Алексеева Н.П.

1) Вычислить медианы среднегодовой температуры в штатах Техас и Калифорния. По критерию Манна-Уитни проверить значимость различия медиан.

	state	temper	Rank
1	CA	59.10	1.00
2	CA	62.10	2.00
3	CA	69.50	3.00
4	CA	71.00	4.00
5	CA	72.60	5.00
6	CA	72.60	6.00
7	CA	73.10	7.00
8	CA	74.30	8.00
9	CA	75.70	9.00
10	CA	77.70	10.00
11	CA	77.90	11.00
12	CA	81.90	12.00
13	TX	82.30	13.00
14	TX	83.50	14.00
15	TX	84.10	15.00
16	TX	84.50	16.00
17	TX	85.00	17.00
18	TX	85.30	18.50
19	TX	85.30	18.50
20	TX	85.90	20.00

Таблица 1: Данные о среднегодовой температуре в штатах Техас и Калифорния

Объемы выборок равны соответственно $n_1 = n_{CA} = 12, n_2 = n_{TX} = 8$. Для вычисления медиан упорядочиваем данные в каждой группе по возрастанию и определяем медиану как среднее арифметическое между элементами, стоящими на 6 и 7 местах в штате Калифорния и на 4 и 5 местах для штата Техас. Получаем медианы 72.85 и 84.75 соответственно.

Расставляем ранги в объединенной выборке (одинаковым значениям присваиваем ранги, равные среднему арифметическому) и суммируем ранги для штата Техас. $R_2 = R_{TX} = 132$. Вычисляем статистику

$$Z = \frac{R_2 - \frac{n_2(n_2+1)}{2} - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} = \frac{132 - \frac{8\cdot 9}{2} - \frac{8\cdot 12}{2}}{\sqrt{\frac{8\cdot 12\cdot 21}{12}}} = 3.7 = Z_*$$

Абсолютное значение статистики Z_* сравниваем с 1.96. Так как $3.7 > 1.96$, то гипотеза однородности отвергается, и различие между медианами считаем значимым. Если бы абсолютное значение статистики Z_* было меньше 1.96, то мы бы считали различие между медианами незначимым.

2) Проверить однородность дыннфх по медианному критерию.

Второй метод проверки однородности заключается в использовании медианного критерия. Вычисляем общую медиану 77.8, затем в каждом штате количество городов, больше общей медианы и меньше общей медианы и собираем результаты в таблицу сопряженности.

```
    CA TX
 <M 10  0
>=M  2  8
```

Для проверки однородности при помощи медианного критерия по таблицы сопряженности вида

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

вычисляется статистика χ^2, равная

$$\chi^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)},$$

где $n = a+b+c+d$. Гипотеза однородности отвергается, если полученное значение статистики $\chi^2_* > 3.84$. В данном случае $\chi^2_* = 13.3 > 3.84$, поэтому различие по медианам нельзя считать случайным.

3) По критерию знаков проверить однородность зависимых выборок. Рассмотрим результаты двух контрольных работ и проверим гипотезу о том, что успеваемость осталась на прежнем уровне.

	$Estimation_1$	$Estimation_2$	$sign(Estimation_2 - Estimation_1)$
1	3	2	-1
2	3	3	0
3	4	5	1
4	4	5	1
5	3	4	1
6	2	4	1
7	2	3	1
8	2	2	0
9	2	5	1
10	3	5	1
11	2	4	1

Вычисляем знаки разностей оценок, обозначаем через n количество ненулевых разностей, $n = 9$, из них $n_- = 1, n_+ = 8$. Воспользуемся таблицей критических значений.

	n	N_
1	7	1
2	8	2
3	9	2
4	10	2
5	11	3
6	12	3
7	13	4
8	14	4
9	15	4
10	16	5

Таблица 2: Критические значения для числа минусовых разностей.

Для уровня значимость 0.05 при $n = 9$ критическим значением являются две минусовые разности, $n_- = 1 < 2 = N_{0.05}(9)$, поэтому можно считать, что имеет место значимое улучшение успеваемости.