Задания к лекции 3. Анализ нечисловых данных.

Вариант 1

Задача 1. Вычислить меры сходства Серенсена и Жаккара для сравниваемых наборов ['Г', 'А', 'Б', 'И', 'Т', 'У', 'С'] и ['О', 'Д', 'У', 'В', 'А', 'Н', 'Ч', 'И', 'К'] без учета порядка входящих букв.

Задача 2. Сравниваются флористические списки A и B. Если к списку A добавить несколько новых, не содержащихся в B видов, как поведет себя мера Жаккара (увеличится, уменьшится, останется неизменной)?

Задача 3. Показать, что в параметризации Б.И. Семкина $K_{0;-1}$ – совпадает с выражением для коэффициента Серенсена.

$$K_{\tau;\eta} = \left(\frac{K_\tau^\eta(A,B) + K_\tau^\eta(B,A)}{2}\right)^{1/\eta},$$
$$K_\tau(A,B) = \frac{|A \cap B|}{(1+\tau)|A| - \tau|A \cap B|},$$
$$K_\tau(B,A) = \frac{|A \cap B|}{(1+\tau)|B| - \tau|A \cap B|}$$
$$-1 < \tau < \infty, -\infty < \eta < \infty$$

Задача 4. Зависит ли скорость изменения коэффициента Жаккара при увеличении числа общих элементов двух сравниваемых множеств, если число элементов объединения этих множеств в процессе изменения остается постоянной.

Задача 5. Даны два мультимножества, состоящие из символов:

$$A = [1, 1, 2, 2, 2, 4, 7, 8, 9, p, p]$$
$$B = [4, 5, 2, 2, 2, 1, 4, 4, 4, 4, q, q]$$

Вычислить меры Жаккара, Дайса и Кульчинского.

Задача 6. Вычислить меру Дайса для двух списков (с повторениями) видов `lec3_measures_var1.dat`.

Caldesia reniformis;	Caldesia reniformis
Caldesia reniformis;	Strobilanthes isophyllus
Strobilanthes isophyllus;	Onychium japonicum
Pseuderanthemum atropurpureum;	Agave filifera
Caldesia reniformis;	Allium spirale

...

Задача 7. Вы сравниваете флористические списки. Матрица мер сходства Жаккара успешно посчитана до вас, но содержит ошибки. Попробуйте определить при сравнении каких именно списков (они нумеруются по строкам/столбцам матрицы) была допущена ошибка. Матрица дана в файле `lec3_bigmat_var1.dat`

Задача 8. Исследуется эффективность действия удобрения на рост растений. Для этого 100 тестовых образцов были разделены в пропорции 2:3 на тестовые (не подвергавшиеся обработки удобрением) и остальные, которые подлежали обработке. Через месяц был произведен учет растений, в результате которого были подсчитаны растения показавшие 10 сантиметровый прирост.

	не удобрено	удобрено	$\sum$
прирост>10 см	10	43	53
прирост<10 см	30	17	47
$\sum$	40	60	100

Можно ли сказать, что данные результаты не следствие случая, а действия удобрения?

Задача 9. Чему равна вероятность наблюдать таблицу сопряженности 2×2 следующего вида:

7	50
45	38

Привести точное выражение для вероятности и вычислить приближенное значение (для вычислений можно использовать среду статистическго анализа R или Python).

Задача 10. Для таблицы сопряженности из предыдущей задачи применить критерий χ^2 (при уровне значимости 0.03) с целью исследования зависимости признаков (для вычислений можно использовать среду статистическго анализа R или Python).

Задача 11. Граф задан матрицей инциденций. Найти его матрицу смежности.

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{matrix}$$

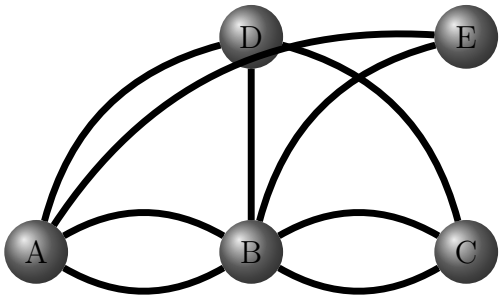
Задача 12. В результате исследования 100 пробных площадей были вычислены их попарные меры сходства. Сходство между пробными площадями считали существенным, если между i-й и j-й площадками вычисленный коэффициент был больше заданного порога. По результатам анализа всех комбинаций была построена матрица смежности: если сходство между i и j было существенным, то соответствующий (i, j)-элемент матрицы выбирался равным 1, в противном случае 0. Найти все совокупности сходных групп пробных площадей, являющиеся компонентами связности графа заданного построенной матрицей смежности. Матрица смежности дана в файле: `lec3_datamat_var1.dat`

Задача 13. Граф задан матрицей смежности. Найти его матрицу инциденций.

$$\begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

Задача 14. Чему равна сумма элементов матрицы инциденций для полного графа с числом вершин n?

Задача 15. Чему равна вероятность реализации графа



Если известно, что у графа обязательно должно быть 9 ребер.

Задания к лекции 3. Анализ нечисловых данных.

Вариант 2

Задача 1. Вычислить меры сходства Серенсена и Жаккара для сравниваемых наборов ['М', 'О', 'Р', 'К', 'О', 'В', 'К', 'А'] и ['М', 'А', 'Р', 'Т', 'Ы', 'Ш', 'К', 'А'] с учетом порядка входящих букв.

Задача 2. Сравниваются флористические списки A и B. Если к списку A добавить несколько новых, не содержащихся в списке B видов, как поведет себя мера Серенсена-Дайса (увеличится, уменьшится, останется неизменной)?

Задача 3. Показать, что в параметризации Б.И. Семкина $K_{1;-1}$ – совпадает с выражением для коэффициента Жаккара.

$$K_{\tau;\eta} = \left(\frac{K_\tau^\eta(A,B) + K_\tau^\eta(B,A)}{2}\right)^{1/\eta},$$
$$K_\tau(A,B) = \frac{|A \cap B|}{(1+\tau)|A| - \tau|A \cap B|},$$
$$K_\tau(B,A) = \frac{|A \cap B|}{(1+\tau)|B| - \tau|A \cap B|}$$
$$-1 < \tau < \infty, -\infty < \eta < \infty$$

Задача 4. Зависит ли скорость изменения коэффициента Дайса при увеличении числа общих элементов двух сравниваемых множеств, если число элементов объединения этих множеств в процессе изменения остается постоянной.

Задача 5. Рассматривая две строки как мультимножества, состоящие из символов:

$$A =' jaccard1901',$$
$$B =' dice1948',$$

вычислить меры Жаккара, Дайса и Кульчинского.

Задача 6. Вычислить меру Жаккара для двух списков (с повторениями) видов `lec3_measures_var2.dat`.

Agave stricta;	Lithops aucampiae
Aloe aristata;	Pseuderanthemum atropurpureum
Aloe aristata;	Allium schoenoprasum
Plumeria rubra;	Hosta rectifolia
Allium spirale;	Allium schoenoprasum

...

Задача 7. Вы сравниваете флористические списки. Матрица мер сходства успешно посчитана до вас, но содержит ошибки. Попробуйте определить при сравнении каких именно списков (они нумеруются по строкам/столбцам матрицы) была допущена ошибка. Матрица дана в файле `lec3_bigmat_var2.dat`

Задача 8. Исследуется эффективность действия удобрения на рост растений. Для этого 100 тестовых образцов были разделены в пропорции 1:3 на тестовые (не подвергавшиеся обработки удобрением) и остальные, которые подлежали обработке. Через месяц был произведен учет растений, в результате которого были подсчитаны растения показавшие 10 сантиметровый прирост.

	не удобрено	удобрено	$\sum$
прирост>10 см	6	18	24
прирост<10 см	19	57	76
$\sum$	25	75	100

Можно ли сказать, что данные результаты не следствие случая, а действия удобрения?

Задача 9. Чему равна вероятность наблюдать таблицу сопряженности 2×2 следующего вида:

25	35
4	18

Привести точное выражение для вероятности и вычислить приближенное значение (для вычислений можно использовать среду статистическго анализа R или Python).

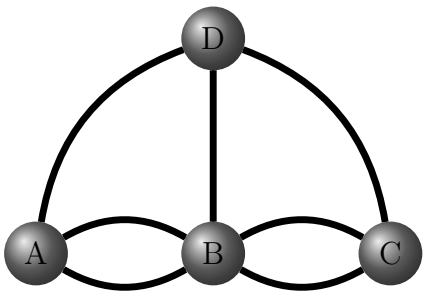
Задача 10. Для таблицы сопряженности из предыдущей задачи применить критерий χ^2 (при уровне значимости 0.02) с целью исследования зависимости признаков (для вычислений можно использовать среду статистическго анализа R или Python).

Задача 11. Граф задан матрицей инциденций. Найти его матрицу смежности.

$$\begin{matrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{matrix}$$

Задача 12. В результате исследования 100 пробных площадей были вычислены их попарные меры сходства. Сходство между пробными площадями считали существенным, если между i-й и j-й площадками вычисленный коэффициент был больше заданного порога. По результатам анализа всех комбинаций была построена матрица смежности: если сходство между i и j было существенным, то соответствующий (i, j)-элемент матрицы выбирался равным 1, в противном случае 0. Найти все совокупности сходных групп пробных площадей, являющиеся компонентами связности графа заданного построенной матрицей смежности. Матрица смежности дана в файле: `lec3_datamat_var2.dat`

Задача 13. Чему равна вероятность реализации графа



Если известно, что у графа обязательно должно быть 7 ребер.

Задача 14. Граф задан матрицей смежности. Найти его матрицу инциденций.

$$\begin{matrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{matrix}$$

Задача 15. Чему равна сумма элементов матрицы смежности для полного графа с числом вершин n?