

## Описание и визуализация распределений признаков

В этом задании вы научитесь исследовать и визуализировать распределения.

### Вы сможете:

- исследовать тип и форму распределения признака, выбрать соответствующие меры центральной тенденции и разброса
- отработать процедуры расчета основных статистик (мер центральной тенденции, мер разброса),
- построить графики, как для простого описания распределений, так и для сравнения
- сформулировать содержательные выводы на основе полученных значений статистик и построенных графиков.

Используя [приложенный файл данных](#), выполните следующие задания:

1. Рассчитайте основные статистики (меры центра и меры разброса) по распределениям всех переменных, имеющихся в файле данных.
2. Выберите наиболее интересный для вас количественный признак и охарактеризуйте его распределение при помощи соответствующих описательных статистик и графиков:
  - 2.1. Какова форма распределения признака?
  - 2.2. Можно ли говорить о том, что распределение признака согласуется с каким-либо теоретическим законом распределения?
  - 2.3. Есть ли в данных «аномалии» / «выбросы»?
  - 2.4. Какие меры центра и вариативности подходят для описания распределений лучше всего? Почему?

При выполнении этой части задания помните, что применимые меры центра и вариативности, а также графики, которые хорошо иллюстрируют распределения, зависят от типа шкалы измерения признака, а также от особенностей / формы распределения.

3. Сравните геодезические и геоинвариантные сферы (экспертная разметка, переменная IsGeo) по выбранному для анализа признаку. Есть ли отличия? В чем они состоят? (для ответа на вопрос используйте статистические и графические инструменты).

При выполнении этой части задания, учтите, пожалуйста, что важно представить все три компонента сравнения: цифры (соответствующие меры центра и разброса), график, иллюстрирующий различия, а также содержательный комментарий: как именно различаются сферы по выбранному признаку.

Вы можете выполнять расчеты в SPSS или в R, работу необходимо загрузить на сайт в формате .doc (.docx) или .pdf.

## **Review criteria**

### **меньше**

Для того чтобы успешно закончить курс, Вам предстоит оценить работы как минимум трех сокурсников. Вы можете оценить и больше работ, если захотите. Тогда и Ваша работа получит больше оценок и обратной связи.

Когда будете оценивать работы, будьте, пожалуйста, конструктивны. Указывайте не только недостатки, но и достоинства работы, предлагайте варианты, как можно было бы улучшить работу (особенно если Ваш комментарий преимущественно критический).

При оценке используйте следующие критерии:

### **Часть 1. Основные статистики по всем переменным файла данных:**

Рассчитаны – 1 балл

Не рассчитаны – 0 баллов

### **Часть 2. Описание распределения признака**

2.1. Есть описание и графическая иллюстрация формы распределения признака

2 – есть и то и другое, описание корректно

1 – отсутствует статистическое или графическое описание, описание распределения в целом корректно

0 – описание распределения отсутствует или некорректно.

2.2. Есть вывод о соответствии распределения какому-либо теоретическому закону

1 – есть корректный обоснованный вывод

0 – нет вывода о соответствии распределения какому-либо теоретическому закону, или вывод есть, но некорректный.

2.3. Описаны ли аномалии и выбросы в данных?

1 – да, аномалии / выбросы описаны

0 – описание аномалий в данных / выбросов отсутствует или ошибочно

2.4. Описаны наиболее адекватные меры вариативности, обоснован выбор мер

2 – меры описаны и обоснованы

1 – меры описаны, но не обоснованы (или в обосновании выбора мер есть ошибки)

0 – описание наиболее адекватных мер центра и вариативности отсутствует или ошибочно.

**Часть 3. Сравнение геодезических и геоинформационных сфер. Сравнение сделано на основе статистик, графиков, содержательного описания различий.**

**3 балла** – полный корректный ответ: есть содержательное, статистическое и графическое сравнение.

**2 балла** – частичный в целом корректный ответ: есть статистики и графики, но нет описания к ним, или описание содержит ошибки; есть статистики и их описание, но нет графика; есть графики и описание, но нет «подкрепления» статистиками (цифрами).

**1 балл** – неполный / частично некорректный ответ: есть ошибки в описании различий, отсутствуют два из трех способов сравнения (статистики, графики, содержательное описание)

Максимальный балл: **10 баллов**

Граница для успешного прохождения: **6 баллов**