**Семинар 7-8.**

**ЧАСТЬ 1**

В файле «Кредиты\_2000\_0.tsv» представлены данные о выданных некоторым банком кредитах.

1. Загрузите данные в Excel, обращая внимание на форматы и разделители. Сколько столбцов у Вас получилось?
2. Проанализируйте типы переменных (непрерывные, дискретные, порядковые, номинальные). Для удобства работы имеет смысл сгруппировать столбцы, содержащие числовые и нечисловые переменные, чтобы они шли по порядку (команда «Вырезать – Вставить вырезанные ячейки»).
3. Найдите в данных пропущенные значения и ошибки (неправильные форматы, лишние символы, языковые несоответствия и т. п.). Используйте фильтры. **Исправьте ошибки.**

Работа с выбросами

1. Проведите анализ выбросов в этих данных. С помощью каких инструментов это можно сделать? Применим ли этот анализ к качественным признакам?

Убедитесь, что те признаки, которые Вы собираетесь анализировать, **действительно** являются числовыми!

Подсказка: чтобы проверить, является ли значение числом, попробуйте применить к нему любую математическую операцию (например, прибавить 1)

Подсказка 2: от некоторых ошибок можно избавиться с помощью функций Excel - ПЕЧСИМВ(), ЗНАЧЕН () и т. п.

Подсказка 3: чтобы посчитать количество выбросов используйте функцию СЧЕТЕСЛИ(), предварительно определив граничные значения (используйте функцию «СЦЕП()”)

1. Проанализируйте еще раз признак «Максимальный выданный кредит». Чем может быть обусловлен такой разброс данных? Как можно это исправить?

Работа с пропусками

Помните, что для симметричных распределений центр распределения одинаково хорошо характеризуется и средним, и медианой, а для несимметричных распределений (асимметричность больше 1 по модулю) среднее значение неинформативно. Мода применяется для качественных признаков. Однако иногда надо заменять пропущенные значения не на основе статистических правил, а руководствуясь здравым смыслом. Предложите и обоснуйте, чем заполнять пропущенные значения в полях:

1. «Кредитный̆ рейтинг».
2. «Годовой̆ доход».
3. «Стаж на текущем месте работы».
4. «Срок с последнего нарушения условий кредитного договора».

**Заполните пропуски.**