Отчёт по лабораторной работе №1

Тема: Структура и описание данных

Задание:

Откройте файл с данными. Приведите его к формату, подходящему для обработки.

1. Выясните структуру данных. Для этого:
2. Разделите переменные на категориальные и количественные
3. Для каждой переменной определите возможные шкалы, выберите шкалу, которая даст возможность получить больше информации.

Для категориальных переменных укажите – номинальными или порядковыми они являются.

Определите, какие качественные переменные могут задавать структуру набора данных. Какие из них будет полезно использовать для фильтрации, разделения выборки и выяснения того, какие зависимости есть в выборке, что может помочь ответить на те вопросы, которые заданы в описании набора данных, выявить, какие исследования можно провести. В вашем исследовании должны быть отражены частоты появлений разных значений качественных переменных; распределения количественных переменных по выборке, распределения, полученные по разным категориям

1. Используя Excel, постройте сводные таблицы для категориальных переменных и таблицы частот для количественных
2. Постройте диаграммы, которые иллюстрируют структуру , для качественных переменных (4-6 штук, с заголовками, подписями осей и пояснениями, что они показывают)
3. Используйте значения качественных переменных как фильтры для количественных
4. Для количественных переменных постройте гистограммы и графики (кумулятивный, частотный, простой, диаграмма рассеяния - какие будут уместны) которые иллюстрируют распределения (4-6 штук, с заголовками, подписями осей и пояснениями, что они показывают)
5. Посчитайте возможные описательные характеристики переменных
6. **Должны быть задействованы в исследовании все переменные выборки**
7. Сформулируйте, какие задачи могут решаться с применением этих данных.  
   Приведите несколько формулировок

Ход работы:

1. Данные из csv файла импортированы в xlsx-файл путём выбора *Данные->Получить->Из текста*. В качестве разделителя выбран символ запятой. Также для числовых значений с плавающей точкой был заменён разделитель на точку.
2. Для каждой переменной была определена шкала и указано номинальная ли она или порядковая, а также качественная или количественная и имеет ли смысл использовать её для фильтра.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Тип | Шкала | Фильтр |
| Sex | Категориальные | Номинальная | Да |
| School Code | Категориальные | Номинальная | Да |
| Playing Often | Категориальные | Порядковая | Да |
| Playing Games | Категориальные | Номинальная | Да |
| Mother Education | Категориальные | Порядковая |  |
| Father Education | Категориальные | Порядковая |  |
| Parent Revenue | Категориальные | Порядковая |  |
| Playing Years | Количественные | Относительная | Да |
| Playing Hours | Количественные | Относительная | Да |
| Grade | Количественные | Относительная |  |

1. Для общего исследования сформулирована тема: «Как влияет воспитание ребёнка на его времяпрепровождение».
2. Для исследования были выбраны следующие сводные таблицы и определены следующие вопросы:

* Пол/Количество – какое соотношение пола в данных?
* Образование матери, отца/Оценка – влияет ли образование матери и отца на образование ребёнка?
* Образование матери, отца/Время игры – влияет ли образование матери и отца на увлечённость развлечениями ребёнка?
* Достаток родителей/Время игры – влияет ли достаток родителей на увлечённость развлечениями ребёнка?
* Время игры/Оценка – влияет ли количество времени, затраченное на развлечения на образование?
* Пол/Время игры – влияет ли разница в воспитании мальчика от девочки на их увлечённость видеоиграми?
* Школа/Время игры – влияет ли школьная дисциплина на то, сколько времени ученик тратит на развлечения?
* Игровой стаж/Время игры – Играют ли старые игроки больше, чем новички?

1. По сводным таблицам построены сводные диаграммы. На основе которых были даны ответы на поставленные выше вопросы:

* Соотношение пола в данных оказалось 50% (386)/50% (384);
* Образование отца и матери напрямую влияет на успеваемость ребёнка (значение 0 учитывать не стоит т.к. их слишком мало):
* Образование родителей слабо влияет на игровое время:

* Достаток родителей напрямую влияет на игровое время:
* Время, проводимое за играми, обратно влияет на успеваемость ученика (чем больше игрового времени, тем меньше успеваемость):
* Один из полов в день тратит меньше времени на игры, чем другой:
* В разных школах игровое время различно (о причине судить сложно):
* Игроки с большим стажем больше увлечены играми: