

Лабораторная работа 2. Описательный и визуальный анализ данных

Задания

Базовая графика

1. Загрузить набор данных NCbirths из пакета Stat2Data и преобразовать в факторные переменные следующие поля (в скобках указаны уровни фактора, которые необходимо задать):
 - Plural (1,2,3)
 - Sex (male, female)
 - Marital (yes, no)
 - RaceMom (white, black, american indian, chinese,japanese, hawaiian, filipino, other asian or pacific islander)
 - HispMom (Cuban, Mexican, not Hispanic, Other Hispanic, Puerto Rico, Central/South America)
 - Smoke (no, yes)
 - Low (no, yes)
 - Premie (no, yes)

Замечание: Прочитать информацию по набору данных можно, выделив название набора и нажав F1

2. С помощью функции *summary()* вывести информацию о следующих численных характеристиках признаков - минимальные и максимальные значения, значения квартилей - для количественных переменных, число объектов каждой категории - для категориальной переменной.

На основе полученной информации уметь описать набор данных.

3. С помощью функции *plot()* построить график зависимости:
 - для нечетных вариантов: веса новорожденного от продолжительности беременности
 - для четных вариантов: веса новорожденного от возраста матери

Подписать график и названия осей. Цвет маркеров выбрать любым, отличным от черного.

4. Построить гистограмму:
 - Для нечетных вариантов: распределения веса новорожденных, матери которых относятся к белой расе и курят (количество интервалов группировки установить равным 10)

- **Для четных вариантов:** распределения веса новорожденных, матери которых относятся к белой расе и курят (количество интервалов группировки установить равным 25)

При этом по оси ординат выводить *относительные* частоты встречаемости значений из интервалов группировки, цвет столбцов установить любым, отличным от черного, подписать график и оси

На той же координатной плоскости построить график плотности вероятности. Сделать вывод о характере распределения.

5. Построить диаграммы размаха ("ящики с усами")

- **Для нечетных вариантов:** распределения веса новорожденных для всех значений факторной переменной MomRace;
- **Для четных вариантов:** распределения веса новорожденных для всех значений факторной переменной Plural;

Подписать график и названия осей.

Интерпретировать полученные диаграммы.

6. Построить столбиковую диаграмму

- **Для нечетных вариантов:** среднего веса новорожденных при разных значениях факторной переменной MomRace;
- **Для четных вариантов:** среднего веса новорожденных при разных значениях факторной переменной Plural

7. Проанализировать совместное распределение категориальных переменных. Для этого необходимо вывести на экран число сочетаний разных значений предложенных трех признаков, а затем построить мозаичную диаграмму.

- **Для нечетных вариантов:** исследуемые признаки: Sex, Plural, Low;
- **Для четных вариантов:** исследуемые признаки: Smoke, Premie, Low.

8. Построив матричные диаграммы рассеяния с помощью функции pairs.panels пакета psych, проанализировать зависимость между следующими признаками:

- **Для нечетных вариантов:** Plural, BirthWeightGm, Gained;
- **Для четных вариантов:** MomAge, BirthWeightGm, Weeks.

9. Построить диаграмму Кливленда для

- **Для нечетных вариантов:** возраста матерей новорожденных с низким весом, разделив на 2 группы детей с курящими и некурящими матерями, и выделив разными цветами точки данных из этих двух групп;
- **Для четных вариантов:** веса новорожденных из двоен, разделив их по возрасту матерей на 3 группы и выделив разными цветами точки данных из этих трех групп

Все числовые данные по осям должны быть упорядочены, см. пример на рисунке ниже.

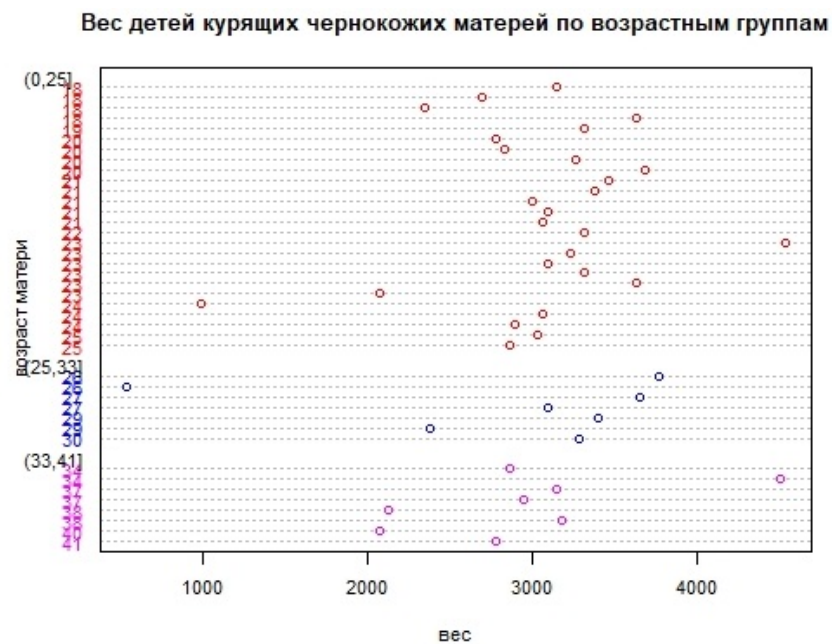


Рис. 1.1: Диаграмма Кливленда

Графика ggplot2

Построить диаграмму с областями (Area Chart),

- **Для нечетных вариантов:** описывающую зависимость числа новорожденных от возраста матерей с разбивкой на группы по расовой принадлежности матери. Использовать цветовую схему отличную от встроенной.
- **Для четных вариантов:** пропорций новорожденных разного пола от веса при рождении (proportional stacked area graph).