

## Маголого «Анализ данных»

Гаврилина Александра,  
магистратура «Суперкомпьютерное моделирование в науке и инженерии»,  
группа МСКМ191

[aagavrilina@edu.hse.ru](mailto:aagavrilina@edu.hse.ru)

[gavrilinasanya@gmail.com](mailto:gavrilinasanya@gmail.com)

### Анализ данных в Stata

Ответьте, пожалуйста, на вопросы:

**1. Определите подходящий тип переменной, задаваемой в Stata для анкетных вопросов:**

- a. Возраст (в годах) int
- b. Пол int
- c. Ежегодный доход компании (в рублях) float
- d. Курс доллара (в рублях) float

**2. Что нужно прописать в командной строке, чтобы создать переменную “age”, отражающую возраст респондента в годах?**

`generate int age = .`

**3. Напишите команду, позволяющую переименовать переменную sex в переменную gender.**

`ren sex gender`

**4. Напишите команду, позволяющую задать переменной age метку «Age in years».**

`label variable age "Age in years"`

**5. Откройте файл nlsw88.dta.**

**a. Сколько уникальных меток присутствует в этой базе данных?**

7

**b. Проведите частотный анализ переменной occupation и ответьте на вопросы ниже:**

- Сколько менеджеров (Managers/admin) участвовали в исследовании?

264

- Каков процент технических работников (Professional/technical) от общего числа людей, принявших участие в исследовании?

14.11%

- Каков процент технических работников (Professional/technical) от числа людей, давших значимые ответы на вопрос о том, к какой категории работников они принадлежат?

14.17%

**с. Подсчитайте указанные ниже статистические характеристики для переменной age:**

- Мода 35
- Медиана 39
- Среднее арифметическое 39.15316
- Размах 12
- Стандартное отклонение 3.060002
- Стандартная ошибка среднего 0.0645679
- Межквартильная широта 6
- Децильное отношение 1.26

**Подсчитайте показатели асимметрии и эксцесса для переменной age. Сделайте выводы о симметричности и пологости распределения значений этой переменной.**

Skewness 0.2003234

При положительной асимметрии в распределении чаще встречаются более низкие значения признака. Величина больше 0, распределение растянуто вправо.

Kurtosis 1.932389

Значение меньше 3 => плосковершинное распределение

**d. Подсчитайте среднее значение переменной age для респондентов состоящих и не состоящих в браке. Какую команду вы использовали для получения результата (см. запись в командной строке)?**

```
tabstat age, statistics( mean ) by(married)
```

```
single 39.21891
```

```
married 39.1165
```

**е. Укажите максимальный возраст респондентов, не состоящих в браке и имеющих значение переменной tenure больше 50. Какую команду вы использовали для получения результата (см. запись в командной строке)?**

```
tabstat age if tenure > 50 & married == 0, statistics( max ) by(married)
```

```
45
```

**f. Анализируя переменные collgrad и married подсчитайте:**

- сколько респондентов, являющихся выпускниками колледжа, состоят в браке?

```
tabstat collgrad if collgrad == 1 & married == 1, statistics( count )
```

```
344
```

- какой процент респондентов, являющихся выпускниками колледжа, состоят в браке?

```
tab collgrad married, row
```

```
64.66%
```

**g. Сколько респондентов имеют следующие характеристики:**

- не состоят в браке (переменная married),
- расовая принадлежность – белые (переменная race),
- не являются выпускниками колледжа (collgrad),
- и не состоят в профсоюзе (переменная union).

## **215**

**Что нужно прописать в командной строке, чтобы получить искомую цифру с использованием команды table?**

```
tabstat idcode if married == 0 & race == 1 & collgrad == 0 & union == 0, statistics( count )
```

**Какой средний возраст и стандартное отклонение по возрасту у группы респондентов, имеющих перечисленные характеристики?**

```
tabstat age if married == 0 & race == 1 & collgrad == 0 & union == 0, statistics( mean, sd )
```

Средний возраст: mean 39.50698

Стандартное отклонение: sd 2.99512

**Какая доля (в процентах) респондентов из этой группы проживает на юге (переменная south)?**

```
tab south if married == 0 & race == 1 & collgrad == 0 & union == 0
31.63%
```