



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Data Analysis National Olympiad - DANO

ТИПЫ ПЕРЕМЕННЫХ. В ЧЕМ ИХ РАЗЛИЧИЕ?

Елена Семерикова,
Валерия Стефаненко

Москва, 2021



ЯНДЕКС.GO:

данные по поездкам в такси



Год	Сезон	Время суток	В пути, мин	Расстояние, км	Стоимость, руб	Стоимость/км, руб	Температура, С	Час пик
2020	осень	вечер	26	23	410	17,8	+5	Нет
2020	осень	день	42	23	654	28,4	+9	Да
2020	осень	утро	27	17	339	19,9	+10	Нет
2020	осень	утро	32	19	472	24,8	+11	Нет
2020	весна	утро	37	32	639	20,0	+7	Нет
2020	зима	день	46	8,5	378	44,5	+1	Нет
2020	зима	день	76	58	1362	23,5	+2	Да
2020	зима	утро	56	27	789	29,2	+1	Нет
2020	зима	утро	69	30	755	25,2	-2	Нет
2019	лето	день	54	27	1016	37,6	+22	Нет
2019	лето	день	6	4	199	49,8	+13	Нет
2019	лето	утро	68	58	1623	28,0	+29	Да
2019	весна	день	20	8	260	32,5	+7	Нет
2019	весна	вечер	19	8	300	37,5	+4	Нет
2019	весна	вечер	37	39	801	20,5	0	Нет
2019	весна	день	25	16	477	29,8	0	Да

КАЧЕСТВЕННЫЕ переменные



Сезон	Время суток	Час пик	Балл пробки
осень	вечер	Нет	3
осень	день	Да	7
осень	утро	Нет	4
осень	утро	Нет	3
весна	утро	Нет	2
зима	день	Нет	2
зима	день	Да	8
зима	утро	Нет	4
зима	утро	Нет	5
лето	день	Нет	3
лето	день	Нет	3
лето	утро	Да	9
весна	день	Нет	3
весна	вечер	Нет	2
весна	вечер	Нет	1
весна	день	Да	7

НОМИНАЛЬНЫЕ

ПОРЯДКОВЫЕ

В Москве, Санкт-Петербурге и других крупных городах сервис Яндекс.Пробки оценивает ситуацию по 10-балльной шкале (где 0 баллов — свободное движение, а 10 баллов — город «стоит»).

КАЧЕСТВЕННЫЕ

переменные



Как определяются категория «время суток»?

Яндекс.Погода ориентируется на социальные сутки (24 часа)

- Утро – с 06:00 до 12:00 часов (часть суток после пробуждения)
- День – с 12:00 до 18:00 часов (пик рабочей активности)
- Вечер – с 18:00 до 00:00 часов (время отдыха после работы)
- Ночь – с 00:00 до 06:00 часов (время сна)

Западные страны ориентируются на AM и PM (по 12 часов)

- AM – с 01:00 до 11:59 часов (дня): morning
- PM – с 12:00 до 11:59 часов (вечера): noon, afternoon, evening, night, midnight

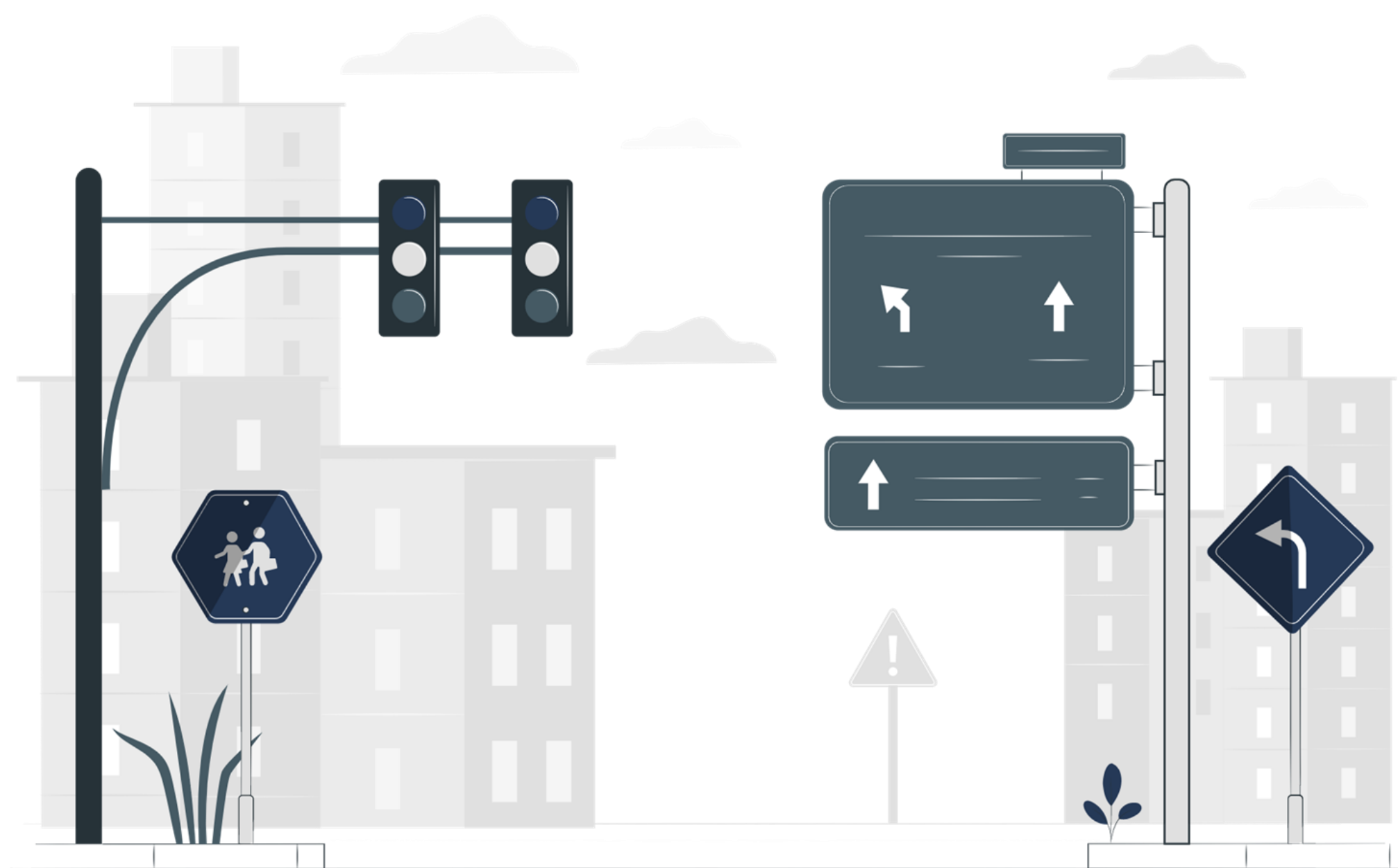
КАЧЕСТВЕННЫЕ переменные

Как определяется час пик?

Час пик по баллу пробки в Яндекс.Пробки

В Санкт-Петербурге при **6** баллах водитель потеряет примерно столько же времени,
сколько в Москве уже при **5** баллах.

То, что в Москве — небольшое затруднение, в другом городе — уже серьёзная пробка.



КОЛИЧЕСТВЕННЫЕ

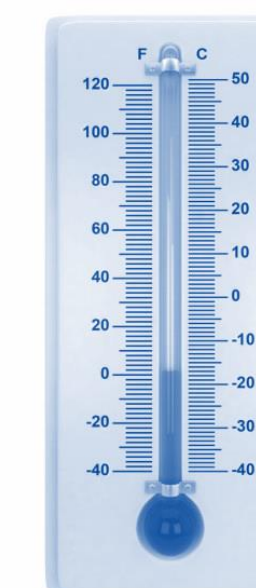
переменные



В пути, мин	Стоимость, руб	Расстояние, км	Температура, С
26	410	23	+5
42	654	23	+9
27	339	17	+10
32	472	19	+11
37	639	32	+7
46	378	8,5	+1
76	1362	58	+2
56	789	27	+1
69	755	30	-2
54	1016	27	+22
6	199	4	+13
68	1623	58	+29
20	260	8	+7
19	300	8	+4
37	801	39	0
25	477	16	0

ОТНОСИТЕЛЬНЫЕ

ИНТЕРВАЛЬНЫЕ



ДИСКРЕТНЫЕ

переменные



**Качественные
номинальные
переменные**
– мальчики и девочки

**Количественные
относительные
переменные**
– количество кубиков
– количество машин
– количество книг на полке

НЕПРЕРЫВНЫЕ

переменные

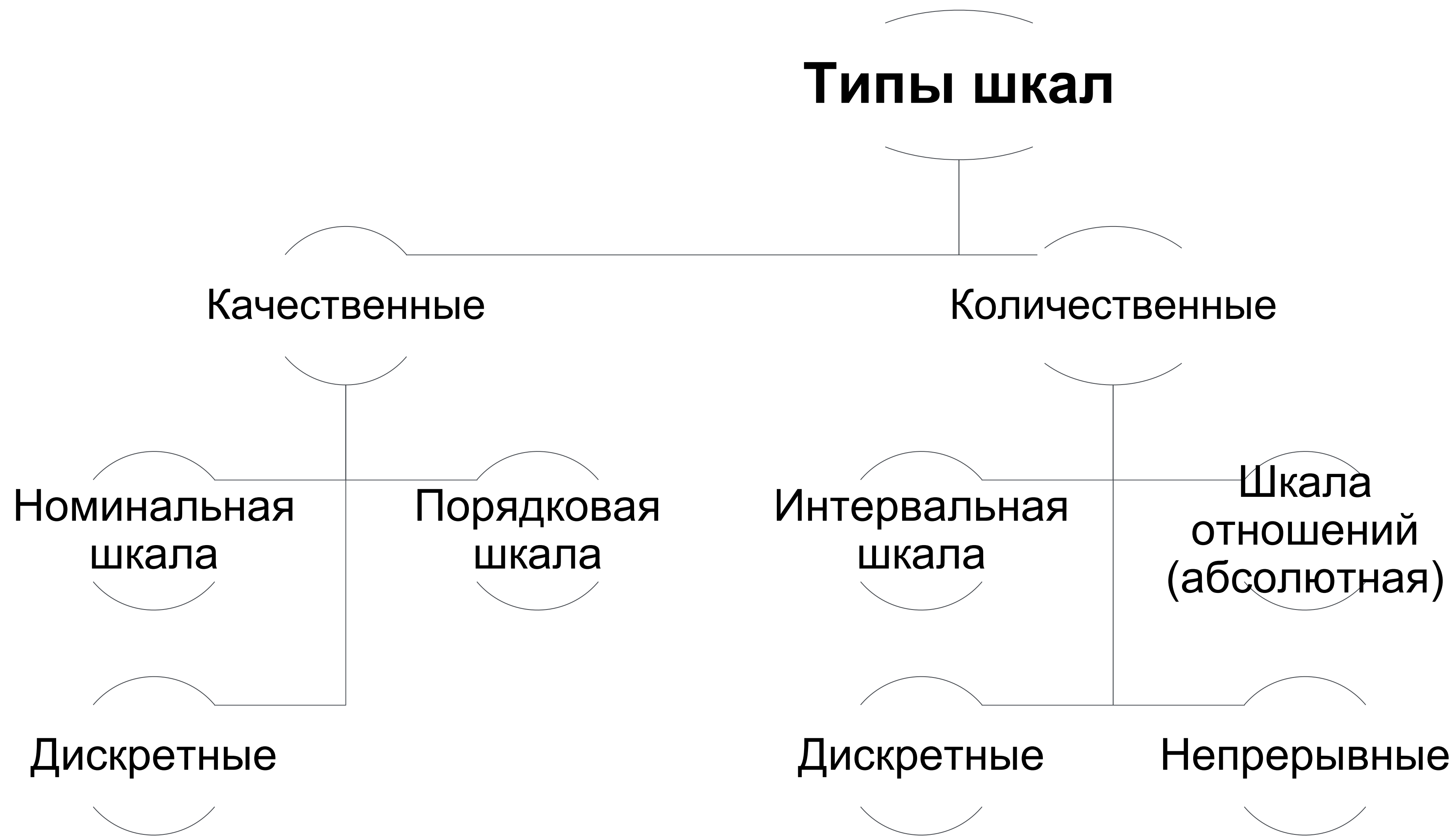


**Количественные
относительные
переменные**
– высота прыжка
– размер дистанции
– скорость
– частота пульса





Типы шкал





ЗАДАНИЕ 1

для самопроверки

Ниже представлены данные по автомобилям. Дайте определения переменным, а именно:

- качественные или количественные;
- дискретные или непрерывные;
- какой тип шкалы измерения: относительная, интервальная, номинальная, порядковая

Расход топлива, км/галлон	Кол-во цилиндров	Мощность	Вес, фунты	Разгон, сек	Год выпуска	Бренд
14,00	8	165	4209	12	1972	US.
31,90	4	71	1925	14	1980	Europe.
17,00	8	140	3449	11	1971	US.
15,00	8	150	3761	10	1971	US.
23,00	8	125	3900	17	1980	US.
13,00	8	158	4363	13	1974	US.
14,00	8	215	4312	9	1971	US.
37,70	4	62	2050	17	1982	Japan.
34,00	4	70	2245	17	1983	Japan.
34,30	4	78	2188	16	1981	Europe.
16,00	8	140	4141	14	1975	US.
11,00	8	180	3664	11	1974	US.
16,00	8	150	3433	12	1971	US.
24,00	4	95	2278	16	1973	Japan.
24,00	4	90	2430	15	1971	Europe.



ЗАДАНИЕ 2

для самопроверки

Охарактеризуйте представленные переменные с помощью подходящей меры центра: для каждой переменной необходимо посчитать моду или среднее.

Расход топлива, км/галлон	Кол-во цилиндров	Мощность	Вес, фунты	Разгон, сек	Год выпуска	Бренд
14,00	8	165	4209	12	1972	US.
31,90	4	71	1925	14	1980	Europe.
17,00	8	140	3449	11	1971	US.
15,00	8	150	3761	10	1971	US.
23,00	8	125	3900	17	1980	US.
13,00	8	158	4363	13	1974	US.
14,00	8	215	4312	9	1971	US.
37,70	4	62	2050	17	1982	Japan.
34,00	4	70	2245	17	1983	Japan.
34,30	4	78	2188	16	1981	Europe.
16,00	8	140	4141	14	1975	US.
11,00	8	180	3664	11	1974	US.
16,00	8	150	3433	12	1971	US.
24,00	4	95	2278	16	1973	Japan.
24,00	4	90	2430	15	1971	Europe.



ЗАДАНИЕ 3

для самопроверки

1. Найдите описание индекса Index of Economic Freedom
2. Опишите, как устроен индекс. Какие показатели в него входят?



ПОЛЕЗНЫЕ ССЫЛКИ

для изучения

Введение в данные. 2.1 Типы данных – курс Новосибирского государственного университета на Coursera

<https://www.coursera.org/lecture/vvedeniye-dannyye/2-1-tipy-dannykh-WusvK>

Типы шкал (Type of scale)

<https://wiki.loginom.ru/articles/scale-type.html>

Базы данных для тренировки классификации переменных:

<https://databank.worldbank.org/home.aspx>

<https://www.kaggle.com/datasets>



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ