



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Data Analysis National Olympiad – DANO

# КАТЕГОРИАЛЬНЫЕ ПЕРЕМЕННЫЕ

Ответы к заданиям для самостоятельного выполнения



# ЗАДАНИЕ 1

## для самопроверки

Месяц	Стоимость картофеля за кг, руб
Декабрь	20
Январь	21,5
Февраль	21,5
Март	22
Апрель	35
Май	40
Июнь	40
Июль	25
Август	17
Сентябрь	18
Октябрь	18
Ноябрь	18,5

Допустим, вы исследуете стоимость овощей. Для этого приходите на рынок каждый месяц и записываете их среднюю стоимость. Собранные вами данные представлены в таблице справа.

1. Поясните, как закодировать данные на основании времени года?
2. Во сколько раз картофель стоит дороже весной, чем осенью?



# ОТВЕТ НА ЗАДАНИЕ 1

для самопроверки

Сезон	Period	Period_1	Period_2	Period_3	Period_4	Средняя стоимость картофеля за кг, руб
Зима	1	1	0	0	0	21
Весна	2	0	1	0	0	32,3
Лето	3	0	0	1	0	27,3
Осень	4	0	0	0	1	18,2

$32,3 : 18,2 \approx 2$  – картофель стоит в 2 раза дороже весной, чем осенью



# ЗАДАНИЕ 2

## для самопроверки

В Москве можно увидеть довольно много многоэтажек. Однако в последнее время также пользуется популярностью и малоэтажное строительство. Вы располагаете данными о количестве этажей в нескольких домах, а также о технологии их строительства.

1. Подумайте, сколько категорий можно придумать для группировки данных, по какому принципу?
2. Закодируйте данные по принципу технологии.
3. Сформируйте дамми-переменную, исходя из классификации:
  - а. малоэтажный- 1-4 этажа (с учетом мансарды);
  - б. многоэтажный - 5 этажей и выше

Дом	Кол-во этажей	Технология
1	12	Панельный
2	4	Монолитный
3	7	Панельный
4	3	Кирпичный
5	9	Панельный
6	5	Монолитный
7	3	Кирпичный
8	2	Монолитный



# ОТВЕТ НА ЗАДАНИЕ 2

для самопроверки

Можно потенциально разбить выборку на 8 групп: по принципу кол-ва этажей, по принципу технологии или на основании двумерного признака «кол-во этажей + технология».

Дом	Кол-во этажей	Категория	Дамми-переменная	Технология	Кодировка
1	12	многоэтажная	0	Панельный	3
2	4	малоэтажная	1	Монолитный	2
3	7	многоэтажная	0	Панельный	3
4	3	малоэтажная	1	Кирпичный	1
5	9	многоэтажная	0	Панельный	3
6	5	многоэтажная	0	Монолитный	2
7	3	малоэтажная	1	Кирпичный	1
8	2	малоэтажная	1	Монолитный	2



# ЗАДАНИЕ 3

## для самопроверки

Справа представлены данные прогноза погоды для США и Канады. Однако в странах применяются разные метрические системы для измерения температур: в США - градус Фаренгейта, а в Канаде – градус Цельсия.

Ответьте на вопрос, на сколько средняя температура за неделю в Цельсиях в США отличается от средней температуры за неделю в Канаде?

Дата	Температура в США, °F	Температура в Канаде, °C
26/07/2021	87	27
27/07/2021	90	26
28/07/2021	86	24
29/07/2021	87	25
30/07/2021	86	25
31/07/2021	86	25
01/08/2021	84	23



# ОТВЕТ НА ЗАДАНИЕ 3

для самопроверки

В США на  $5,4^{\circ}\text{C}$  теплее в среднем, чем в Канаде.

Дата	Температура в США, $^{\circ}\text{F}$	Температура в США, $^{\circ}\text{C}$	Температура в Канаде, $^{\circ}\text{C}$
26/07/2021	87	31	27
27/07/2021	90	32	26
28/07/2021	86	30	24
29/07/2021	87	31	25
30/07/2021	86	30	25
31/07/2021	86	30	25
01/08/2021	84	29	23
	Средняя температура, $^{\circ}\text{C}$	30,4	25