Национальная олимпиада по анализу данных DANO. 2 этап (вариант 2)

Во всех заданиях с выбором ответа может быть один или несколько правильных вариантов ответа.

Задания с 1 по 3 оцениваются в 2 балла.

Вопрос 1

Как известно, Израиль является одной из лидирующих стран по темпам вакцинации. По данным на июнь 2021 г. в стране вакцинировано 60% граждан (85% взрослого населения). Однако среди заразившихся в этом же месяце (июне 2021 года), как признали власти Израиля, примерно половина была уже вакцинирована. **Что можно сказать об эффективности вакцины на основании этих данных?**

1) Данные не свидетельствуют об эффективности вакцины, т. к. вероятность заразиться составляет 50%, независимо от того, вакцинировался человек или нет
2) Данные не свидетельствуют об эффективности вакцины, т. к. среди вакцинированных есть заразившиеся
3) Данные свидетельствуют об эффективности вакцины, т. к. если бы она не работала, доля вакцинированных среди заболевших была бы равна доле вакцинированных среди всего населения страны
4) Данные свидетельствуют об эффективности вакцины, т. к. вакцинированные переносят болезнь в более легкой форме

Ответ: 3

Вопрос 2

Известно, что у рассматриваемой выборки медиана равна 14, а среднее значение числа трехзвездочных ресторанов в трех странах с наибольшими количествами отличается от аналогичного среднего у трех стран с наименьшими количествами в рассматриваемой выборке на 7,(3). **Чему равно количество ресторанов с тремя звездами Мишлен в Японии?**

1) 16
2) 9
3) Нет верного ответа
4) 28
5) 6
6) 6

Ответ: 1

Вопрос 3

Министерство здравоохранения пытается оценить тяжесть ситуации с эпидемией нового заболевания в стране. Эта болезнь характеризуется тем, что симптомы очень похожи на обычную простуду, однако у части людей может протекать совершенно бессимптомно. Также в данной стране все люди очень ответственны и при возникновении симптомов обращаются к врачу. Для этих целей министерство использует число всех зарегистрированных случаев простудных заболеваний. КАКУЮ ОЦЕНКУ

ЗАБОЛЕВАЕМОСТИ НОВОЙ БОЛЕЗНЬЮ ОНИ МОГУТ ПОЛУЧИТЬ С ПОМОЩЬЮ ТАКОЙ СТАТИСТИКИ?

A) Оценку сверху

B) Оценку снизу

C) Точную оценку

D) Нельзя ответить на основе данной информации

Ответ: D

Задания с 4 по 10 оцениваются в 5 баллов

Вопрос 4

Вы хотите изучать влияние исторического распределения школ на текущий уровень человеческого капитала, который включает в себя образование, престижность профессии, продолжительность жизни и другие факторы. **Укажите, какие из графиков способствуют поиску ответа на поставленный вопрос:**

1) Дистанция до ближайшей школы в прошлом ~ текущий уровень грамотности, по индивидам
2) Текущее количество школ в регионе ~ текущий средний уровень образования, по регионам
3) Категория занятости (работники ВУЗов, профессиональные рабочие, низкоквалифицированный труд и др.) ~ дистанция до ближайшей школы в прошлом, по индивидам
4) Уровень образования по годам в изучаемом и соседнем регионе, по регионам
5) Нет верного ответа

Ответ: 1 и 3

Вопрос 5

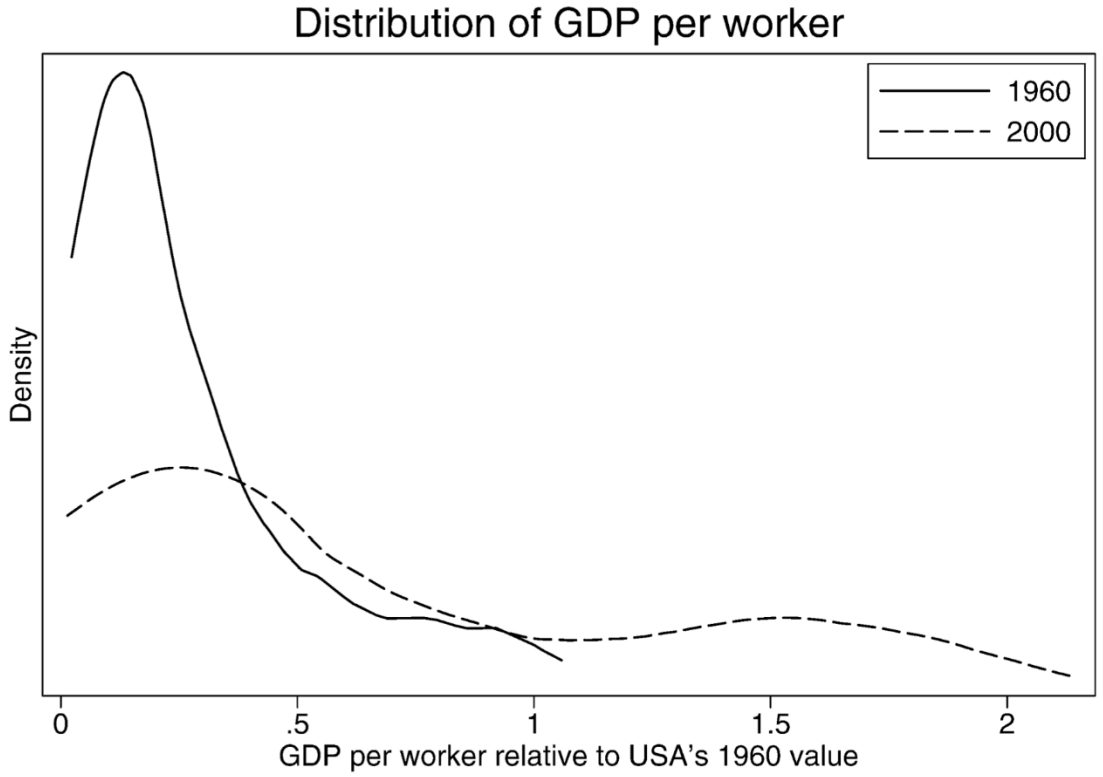
Современные технологии распознавания лиц позволяют правоохранительным органам вычислять преступников на улицах. Алгоритм разделяет людей на обычных граждан и подозреваемых. Подозреваемым считается тот, чью фотографию опознал алгоритм в базе. Предположим, что алгоритм никогда не отмечает преступников как обычных граждан. Известно, что все преступники из базы сейчас находятся на территории города и их разыскивает полиция. **Какие утверждения верны, если все население города попало на камеру за определенный период?**

1) Алгоритм может распознать как подозреваемых не меньше людей, чем есть в базе преступников.
2) Алгоритм распознает столько подозреваемых, сколько в базе преступников.
3) Алгоритм может распознать меньше подозреваемых, чем есть в базе преступников.
4) Алгоритм распознает половину граждан как подозреваемых.
5) Нет верного ответа

Ответ: 1

Вопрос 6

Группа ученых, исследующих экономический рост, посчитала за 1960 и 2000 года отношение ВВП на одного работника в разных странах к этому же показателю для США. На графике ниже представлена функция плотности данного соотношения. **Какие утверждения из приведенных ниже являются верными на основе приведенного графика?**



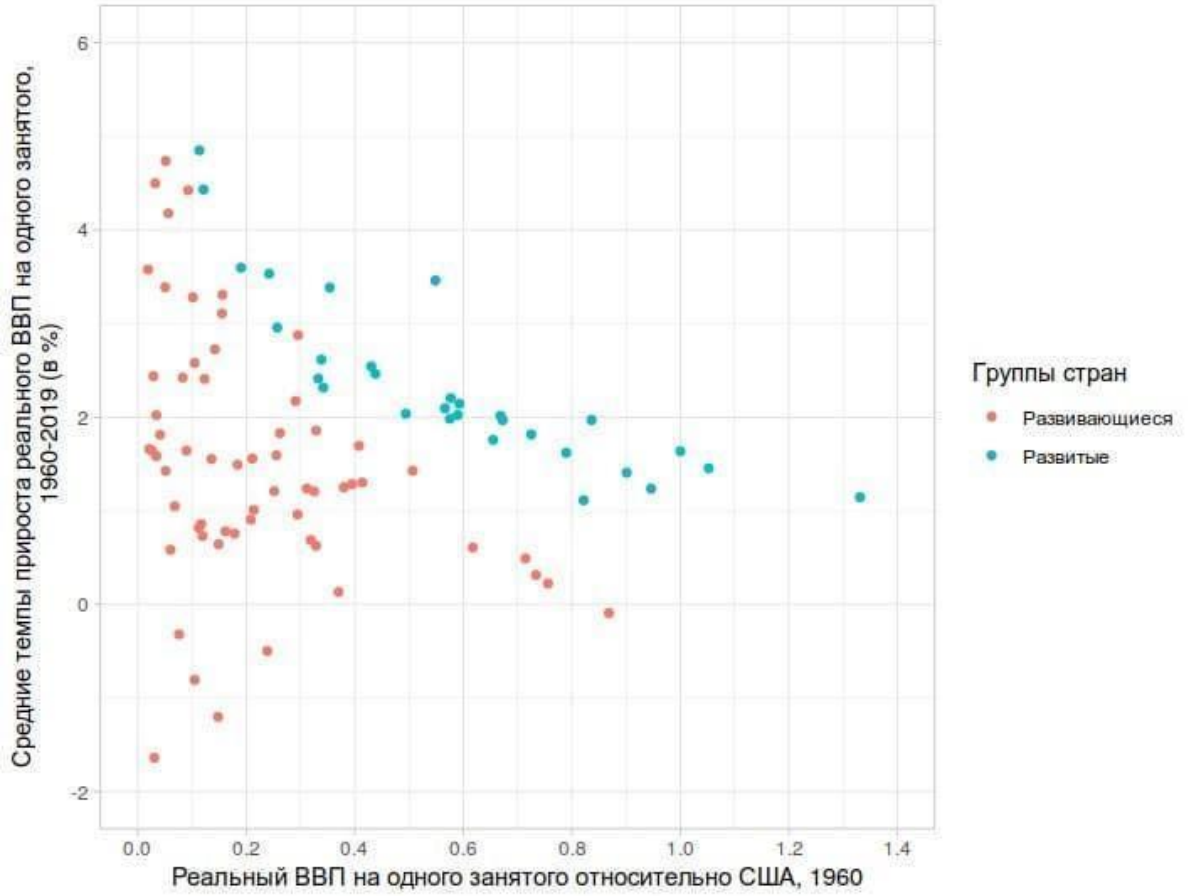
Источник: Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. Growth econometrics.

1) В 2000 году в большинстве стран ВВП на работника был ниже, чем в США
2) ВВП на работника в США вырос за период 1960-2000гг.
3) С 1960 по 2000 год ВВП на работника во всех странах рос медленнее, чем США
4) В 1960 году были страны, ВВП на работника в которых превышал этот показатель для США больше, чем в 2 раза
5) Нет ни одного верного ответа

Ответ: 1

Вопрос 7

Конвергенция в теории экономического роста – это процесс, когда страны с меньшим уровнем доходов на душу населения догоняют страны с более высокими доходами. На графике ниже представлена диаграмма взаимосвязи темпов роста в 1960-2000 годах и реального ВВП на одного работника по отношению к ВВП на одного работника в США в 1960 году. Также на графике разными цветами выделены развитые и развивающиеся страны на основании методологии МВФ по состоянию на 2018 год.



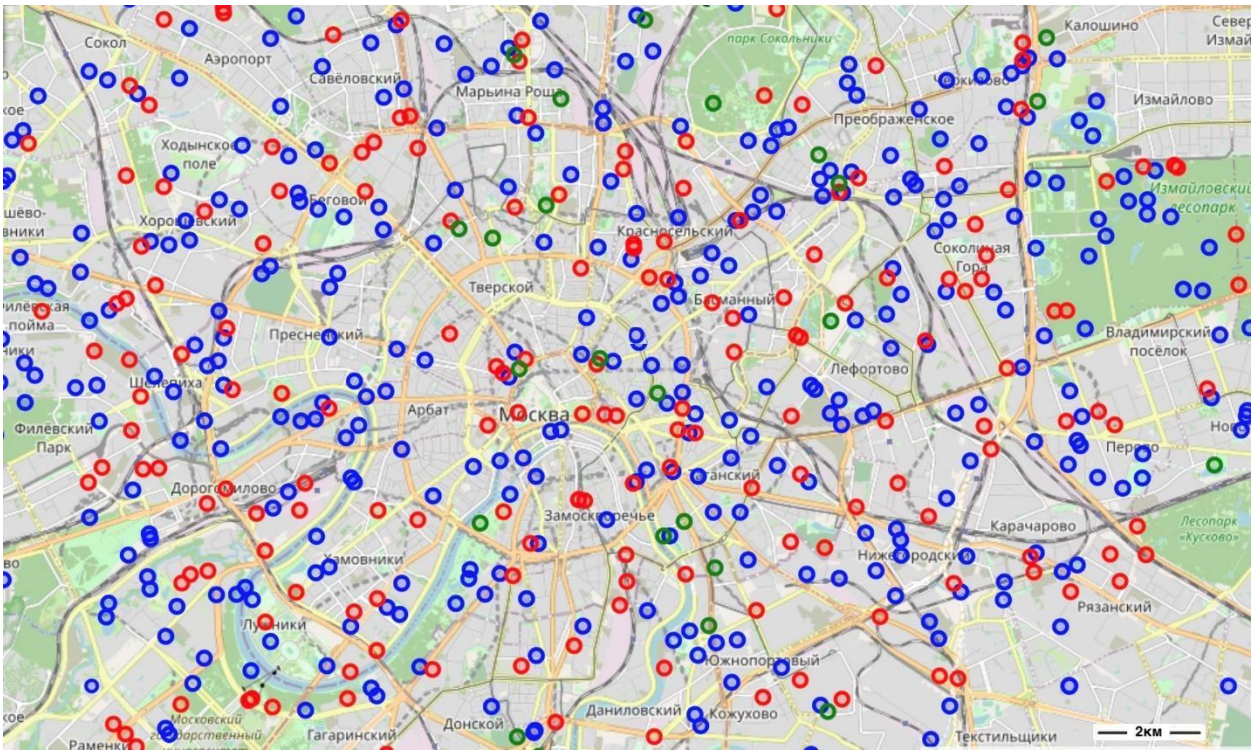
Источник: Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. Growth econometrics, построение авторов

1) Наблюдается конвергенция в уровне доходов на одного работника в рассматриваемый промежуток времени только в развитых странах, поскольку эти точки демонстрируют отрицательную взаимосвязь показателей
2) В развивающихся странах и выпуск на работника, и темпы его роста оказываются ниже, чем в развитых, поэтому конвергенция не наблюдается
3) В течение 60 лет выпуск на одного работника возрастает не во всех странах, а значит конвергенция наблюдается
4) В большинстве развитых темпы роста выпуска на одного работника ниже, чем в США, поэтому конвергенции не наблюдается
5) Нет ни одного верного утверждения

Ответ: 1

Вопрос 8

Предположим, что есть некая интернет-платформа, пункты выдачи заказов которой нанесены на карту по типам. Красные точки отображают собственные пункты выдачи заказов, синие точки отображают партнерские пункты, а зеленые – отделения почты, через которые также могут выдаваться интернет-заказы. **Какие выводы на основе данной карты можно сделать?**



1) компания в основном опирается на собственные пункты выдачи
2) собственные пункты выдачи генерируют большую прибыль
3) отделения Почты распределены неравномерно по городу
4) человек, находящийся в любом месте на данной карте, сможет найти пункт выдачи заказов в радиусе 3 км
5) нет ни одного верного ответа

Ответ: 3

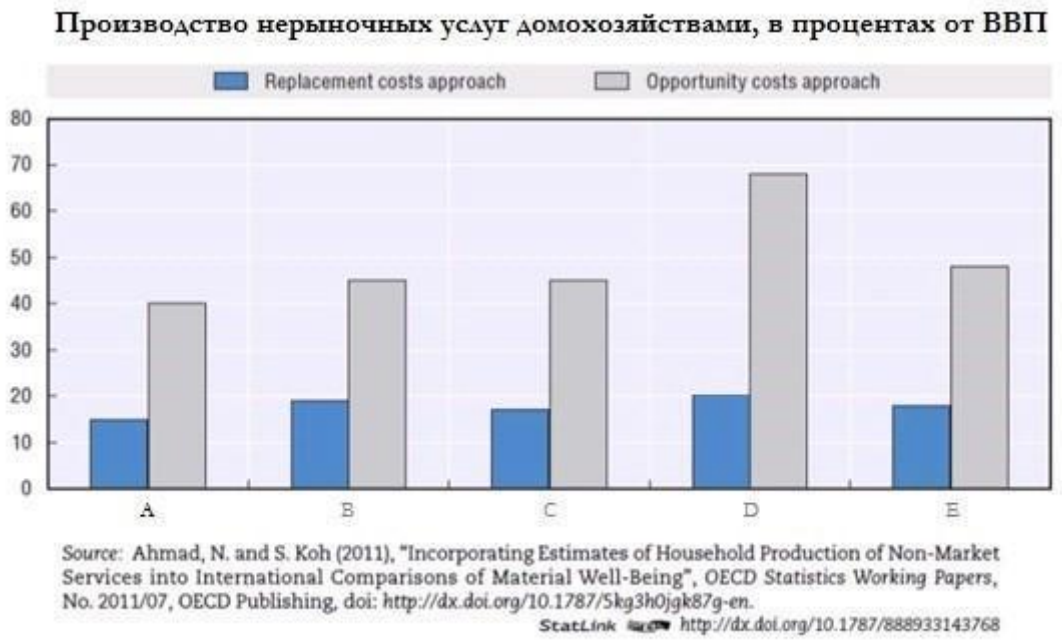
Вопрос 9

Экономисты часто задаются вопросом относительно учета в ВВП различных показателей, которые невозможно посчитать напрямую. Так, например, включили в ВВП проживание в собственной квартире исходя из цены аренды помещения на рынке. Еще один интересный и довольно значимый вклад в выпуск может внести работа людей на самих себя в рамках ведения своих ежедневных дел. Домохозяйка, которая сидит дома и занимается исключительно своими бытовыми заботами, по факту оказывает довольно большое количество услуг: уборщицы, повара, няни и т.п.

Экономисты предлагают два альтернативных способа оценить стоимость оказываемых услуг. Первый способ – replacement costs approach (подход на основе издержек замещения) – это оценить стоимость услуг, которые домохозяйки оказывают своему домохозяйству,

исходя из рыночных цен на аналогичные услуги (то есть сложить сколько стоит нанять няню, повара и горничную для выполнения тех же работ). Второй способ – opportunity costs approach (подход на основе альтернативных издержек) – это оценить, сколько бы в среднем женщина зарабатывала, выйдя на рынок труда. На графике ниже приведены оценки неучтенного ВВП из-за труда домохозяек для разных стран (A, B, C, D и E).

Какие выводы мы можем сделать, основываясь на данных графиках?

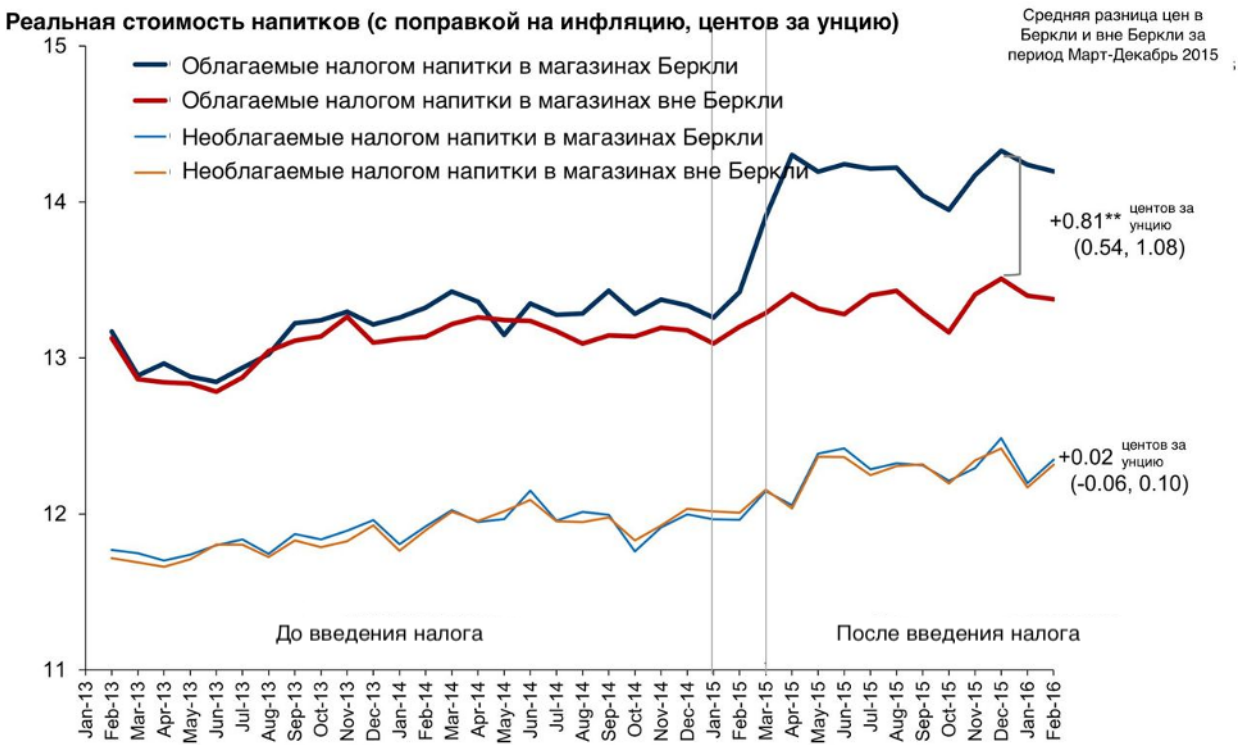


1) В стране A домохозяйки оказывают сами себе меньше услуг, чем в стране E
2) Во всех странах из приведённых на графике зарплаты нянь, горничных и поваров в среднем ниже, чем средние зарплаты в других профессиях
3) Мы ничего не можем сказать о соотношении средних рыночных зарплат в разных странах
4) Разница между среднерыночной зарплатой и зарплатой поваров, горничных и нянь больше в стране D, чем в стране A
5) Нет ни одного верного ответа

Ответ: 2 и 3

Вопрос 10

Напитки, содержащие сахар, считаются вредными для здоровья, и поэтому в Беркли (в США) в начале 2015 года ввели налог на продажу этих напитков для уменьшения их потребления. Первая вертикальная линия на графике показывает момент объявления о введении налога, вторая – момент начала выплат налога продавцами. Однако, чтобы отследить чистый эффект налога, исследователи собрали данные из магазинов в соседних городах, где налог не был введен, и сравнили их с ценами на напитки в Беркли. Цены на сахар в указанный период времени не менялись. Данные представлены в виде графиков цен на напитки в Беркли и в других городах.



Какие свойства графика позволяют нам сделать вывод о воздействии налогов на цену напитков?

1) Цены на облагаемые налогом напитки до введения налога в Беркли и вне Беркли должны меняться по-разному
2) Цены на облагаемые и необлагаемые налогом напитки в Беркли до введения налога должны меняться по-разному
3) Цены на облагаемые налогом напитки после введения налога в Беркли и вне Беркли должны меняться одинаково
4) Цены на необлагаемые налогом напитки после введения налога в Беркли и вне Беркли должны меняться одинаково
5) Нет ни одного верного ответа

Ответ: 4

Задания с 11 по 15 оцениваются в 8 баллов

Вопрос 11

В представленном файле (Задача про футбол) вы можете найти информацию о футбольных матчах Английской премьер-лиги сезона 2017-2018 года. **Из приведенных ниже утверждений выберите все верные.**

1) В домашних матчах команды в среднем забивают меньше голов, чем в выездных матчах
2) В выездном матче больше шансов сыграть вничью, чем выиграть матч
3) Большинство матчей выигрывается в гостевых играх
4) Вариация в количестве забитых мячей в домашних матчах выше, чем в гостевых
5) Нет ни одного верного ответа

Ответ: 4

Вопрос 12

Выберите все верные утверждения на основании анализа данных, предложенных вам в файле (**Задача про машины**).

1) Более старые автомобили продаются в среднем по более высоким ценам
2) Автомобили с большим пробегом продаются в среднем по более низким ценам
3) Более старые автомобили имеют в среднем более высокий пробег
4) Автомобили в хорошем состоянии продаются в среднем примерно на 67% дороже автомобилей в плохом состоянии
5) Нет ни одного верного ответа

Ответ: 2, 3, 4

Вопрос 13

Индекс человеческого развития (ИЧР) — интегральный показатель, рассчитываемый ежегодно для межстранового сравнения и измерения уровня жизни, грамотности, образованности и долголетия как основных характеристик человеческого потенциала исследуемой территории. Он является стандартным инструментом при общем сравнении уровня жизни различных стран и регионов.

(Источник: Википедия)

Необходимо посчитать индексы человеческого развития (ИЧР) для представленных стран в файле (**Задача про ИЧР**) (Настоящая методология расчета ИЧР включает в себя некоторые поправки, которые в данном задании опущены. В связи с этим результаты могут отличаться от опубликованных значений ИЧР). Для вычисления ИЧР выполните следующие пункты:

а) Используя индикатор Ожидаемая продолжительность жизни при рождении (Life expectancy at birth), посчитайте Индекс здоровья по представленной формуле:

$$\text{Индекс} = \frac{\text{значение показателя} - \text{минимальное значение показателя}}{\text{максимальное значение показателя} - \text{минимальное значение показателя}}$$

б) Посчитайте Индекс образования, используя два индикатора:

Средняя продолжительность обучения населения в годах и Ожидаемая продолжительность обучения населения, ещё получающего образование, в годах.

Для каждого индикатора вычислите свой индекс по представленной формуле:

$$\text{Индекс} = \frac{\text{значение показателя} - \text{минимальное значение показателя}}{\text{максимальное значение показателя} - \text{минимальное значение показателя}}$$

в) Посчитайте Индекс доходов, используя показатель натуральный логарифм Валового национального дохода на душу населения (GNI per capita). Для его вычисления необходимо взять натуральный логарифм значений (этот показатель уже посчитан и приведен в файле с данными, Log GNI).

(Натуральный логарифм есть обратная функция для экспоненты. Функция является монотонно возрастающей. Ее часто применяют для преобразования различных экономических переменных, которые характеризуются постоянным ростом (например, для ВВП). Для его вычисления в Excel используется функция LN()).

Индекс считается по формуле:

$$\text{Индекс} = \frac{\text{значение логарифма показателя} - \text{минимальное значение логарифма показателя}}{\text{максимальное значение логарифма показателя} - \text{минимальное значение логарифма показателя}}$$

г) Рассчитайте Индекс человеческого развития, вычислив геометрическое среднее по трем индексам. Округлите значения до 3 знаков после запятой и расположите страны в порядке убывания индекса.
Выберите все верные утверждения в соответствии с Вашим решением.

1) Максимальное значение Индекса образования – единица
2) ИЧР равен нулю в трех странах
3) В Казахстане ИЧР выше, чем в России
4) Наибольший Индекс образования в Австралии
5) Нет ни одного верного варианта ответа

Ответ: 3 и 4

Вопрос 14

В этом задании будет рассмотрено влияние появления нового мусоросжигательного завода на стоимость жилья в одном из городов США. Для этого мы будем использовать данные за два года (до и после строительства завода) о домах, находящихся рядом с заводом, и домах, находящихся далеко от завода. Данные вы можете найти в файле **«Задача про дома»**. Единицей наблюдения в нем является дом.

Отметим, что слухи о том, что в городе будет построен новый мусоросжигательный завод, появились после 1978 года, а строительство началось в 1981 году. Мы будем использовать данные о ценах на дома, проданные в 1978 году (период «до»), а также данные о ценах на дома, проданные в 1981 году (период «после»). Гипотеза заключается в том, что цены на дома, расположенные вблизи мусоросжигательного завода, упадут по сравнению с ценами на более удаленные дома.

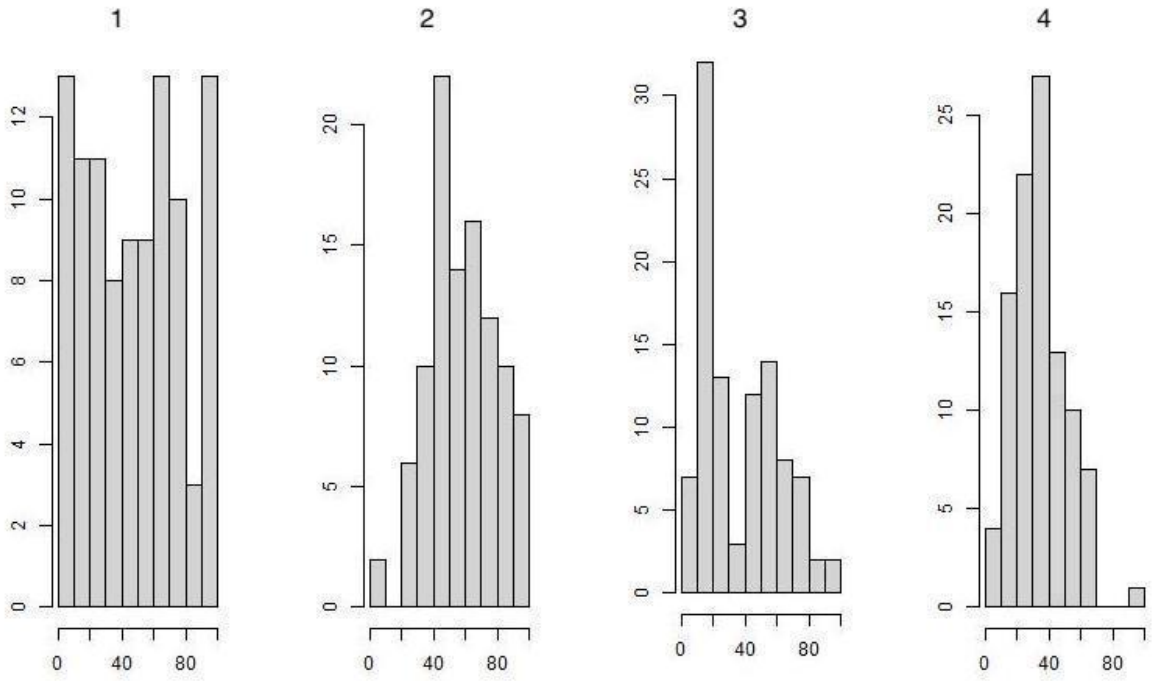
Выберите, какие из представленных ответов соответствуют данным и одновременно помогают обосновать представленную гипотезу.

1) Средняя цена на дома в 1981 году, расположенные далеко от завода, выше по сравнению с ценой домов, расположенных близко к заводу
2) Средняя цена на дома, расположенные далеко от завода, выросла в 1981 году по сравнению с 1978 годом
3) В 1981 году по сравнению с 1978 годом средняя цена на дома, расположенные вблизи к заводу, выросла сильнее, чем средняя цена на дома, расположенные далеко от завода
4) Эффект снижения средней цены на дома в связи со строительством завода составляет примерно 11 тыс.
5) Нет ни одного верного ответа

Ответ: 4

Вопрос 15

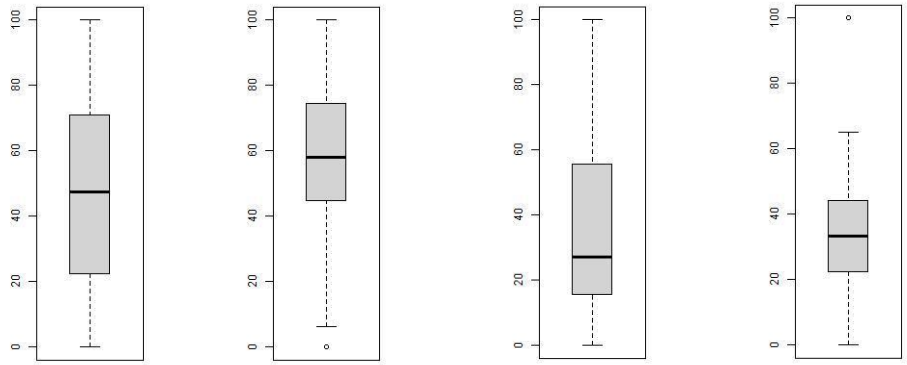
На представленных вам рисунках вы можете найти частотные диаграммы и графики box-plot.
Соотнесите разные типы диаграмм, которые отображают один и тот же набор данных.



В данном задании вы получаете 2 балла за каждое верное сочетание гистограммы и box plot. Неверное сочетание не штрафуется.

Ответ:

1) 2) 3) 4)



Вопрос 16 *(Задание оценивается в 10 баллов)*

В файле «**Задача про студентов**» вам дана таблица, где есть ученики, данные про них, и курсовые проекты, которые они выбрали по приоритетам. Если анкета заполнена

несколько раз, то учитывается последняя отправка. Если два студента хотят выбрать один и тот же проект, то в итоге на него будет записан студент с более высоким рейтингом.

Выберите верное распределение курсовых проектов между студентами, если для каждого нужно выбрать только один проект и на каждый проект может записаться не более одного студента.

В данном задании вы получаете 2 балла за каждое верное сочетание студента и курсового проекта. Неверное сочетание не штрафуется.

Ответ:

1) Современная литература
2) Анализ данных
3) Программирование
4) Микроэкономика
5) Введение в социологию
   История России (лишнее)

Вопрос 17 *(Задание оценивается в 14 баллов)*

Заполните пропуски, чтобы получилось верное рассуждение на тему расчета индекса потребительских цен.

В данном задании вам необходимо выбрать верный вариант заполнения пропуска из предложенных вариантов. Каждый правильно заполненный пропуск оценивается в 2 балла. Неверный ответ не штрафуется.

Инфляция обычно измеряется с помощью индекса потребительских цен. Для расчёта этого показателя берётся фиксированная(ый) [[1]], общая стоимость которой(ого) потом используется для расчёта индекса: стоимость [[5]] в текущий год сравнивается со стоимостью в [[9]] год и это изменение интерпретируется как инфляция.

Однако цены растут неравномерно и из-за эффекта замещения люди, выбирая между двумя очень похожими товарами (разными сортами яблок, к примеру), начинают покупать [[13]] тех, что медленнее дорожают. В результате, фактический состав [[16]] меняется. За счёт этого уровень инфляции, посчитанный по неизменной(ому) [[20]], оказывается [[24]] "настоящего" уровня инфляции.

1) Корзина

5) Корзины

9) базовый

13) больше

16) корзины

20) корзине

24) выше

Вопрос 18 *(Задание оценивается в 8 баллов)*

В представленном файле (**Задача про оценки**) вы можете найти данные о характеристиках различных учеников. **Далее вам предлагается заполнить пропуски в тексте ниже,**

чтобы получить верный текст, который начинающий аналитик представляет коллегам в качестве результатов исследования.

В данном задании вам необходимо выбрать верный вариант заполнения пропуска из предложенных вариантов. Каждый правильно заполненный пропуск оценивается в 2 балла. Неверный ответ не штрафуется.

В своем исследовании мы хотим выяснить, учатся ли более ответственные ученики лучше. Основная гипотеза исследования состоит в том, что [[1]]. Ответственность учащихся мы будем измерять с помощью [[5]]. Проведенные нами расчеты говорят о том, что ответственность и итоговый балл [[6]]. Данный результат [[9]] основную поставленную гипотезу.

1) у более ответственных учеников суммарный балл за все тесты выше

5) переменной test preparation course, потому что более ответственные ученики скорее пройдут подготовительный курс до конца

6) связаны положительно

9) не отвергает

Задания 19 и 20 оцениваются в 1 балл

Вопрос 19

В представленном файле (**Задача про оценки**) вы можете найти данные о характеристиках различных учеников. В этом задании для ответа на вопрос вам будет необходимо посчитать некоторые показатели по данной базе.

Подсказка: для выделения более или менее ответственных учеников используйте наиболее подходящий из следующих трех показателей: lunch, test preparation course или gender.

Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33).

Средний суммарный балл по всем тестам среди более ответственных учеников составляет:

Ответ: 210,19

Вопрос 20

В представленном файле (**Задача про оценки**) вы можете найти данные о характеристиках различных учеников. В этом задании для ответа на вопрос вам будет необходимо посчитать некоторые показатели по данной базе.

Подсказка: для выделения более или менее ответственных учеников используйте наиболее подходящий из следующих трех показателей: lunch, test preparation course или gender.

Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33).

Средний суммарный балл по всем тестам среди менее ответственных учеников составляет:

Ответ: 177,64

*В представленном файле (**Задача про меню**) вы можете найти информацию о различных заказах, сделанных в кафе за месяц. Вас попросили сделать небольшую аналитику по имеющимся продажам. В задании 21-25 вам будет необходимо по очереди заполнить пропуски в предложенном ниже тексте. Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33). Каждое задание оценивается в 2 балла.*

Вопрос 21

Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33).
Всего за указанный период количество заказов составило:

Ответ: 1223

Вопрос 22

Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33).

В среднем в каждом заказе (...) позиции.

Ответ: 2,51

Вопрос 23

Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33).
В среднем в каждом заказе (ответ из предыдущего задания) позиции на общую сумму:

Ответ: 18,77

Вопрос 24

Если ваш ответ представлен дробным числом, запишите его через запятую с округлением до 2 знаков после запятой (например, если вы получили 1/3, то в ответе необходимо указать 0,33).
Всего за указанный период было заказано (...) уникальных позиций из нашего меню.

Ответ: 1365

Вопрос 25

В качестве ответа внесите в поле нужный item_name, представленный словом или словосочетанием, латинскими буквами, в той же форме, в которой оно написано в базе данных, которую вы анализируете.

Из всех видов позиций "item_name" наибольшую выручку в нашем кафе принесла:

Ответ: chicken bowl