

Домашнее задание по курсу

«Теория вероятностей и математическая статистика»

Домашнее задание присылается на адрес dz.tvims@yandex.ru до **23:59 10.06.2023** в виде архива, содержащего следующие допустимые типы файлов:

- Таблица Excel с расчетами (*.xls, *.xlsx, *.xlsm)
- Код из R или python (*.r, *.py)
- Ноутбук Jupyter (*.ipynb)
- Презентация с описанием полученных результатов (*.pptx, *.pdf).

Название архива и название письма должны соответствовать формату:

Номер группы_Фамилия_Excel_R_python.

Пример темы письма: 160_Иванов_Excel_R

Пример названия архива: 160_Иванов_Excel_R.zip

Между словами должны быть нижние подчеркивания. Слово БЭК не используется. В последней части названия должны быть перечислены все программные продукты из набора Excel, R, python, которые использовались в домашнем задании.

Домашние задания, присланные на любые другие адреса электронной почты или оформленные без учета написанного выше, НЕ ПРОВЕРЯЮТСЯ.

Для решения любой задачи может использоваться Excel, R или python.

При использовании R и python могут быть использованы любые библиотеки. Вспомогательные материалы для выполнения задания в Excel можно найти здесь:

<https://www.youtube.com/watch?v=qTRs0Oy3nII>

<https://www.youtube.com/watch?v=4oxsH95dVSI>

Проверяющий оставляет за собой право задать студенту, сдавшему задание в R или python, вопрос по коду. В случае невразумительного ответа баллы за задание снимаются.

В презентации должны быть представлены полученные выводы (кроме пункта 2.1), результаты расчетов (в виде таблиц или графиков) и указанные в задании диаграммы.

В домашнем задании используются следующие параметры: a_1, a_2, a_3 - номер первой, второй и третьей буквы фамилии в алфавите соответственно, b_1, b_2, b_3 - номер первой, второй и третьей буквы имени в алфавите соответственно, c_1, c_2, c_3 - номер первой, второй и третьей буквы отчества (или снова имени, если таковое отсутствует) в алфавите соответственно. За основу используется алфавит, представленный здесь:

http://ru.wikipedia.org/wiki/Русский_алфавит

Если вторая или третья буква в фамилии, имени или отчества отсутствует, то вместо нее используется предыдущая буква.

Задание 1. Базовые описательные статистики

1.1. Основные характеристики выборки (20 баллов)

1. Создайте выборку объемом $n = 100$ из равномерного на $[-a_1, a_2]$ распределения
2. Создайте выборку объемом $n=100$ из нормального распределения с параметрами $\mu = b_1$ и $\sigma^2 = b_2$.
3. Создайте выборку объемом $n=100$ из показательного распределения с параметром $\lambda = 1/c_1$.
4. Для построенных выборок найдите значения:
 - Первого и последнего члена вариационного ряда
 - Выборочного среднего
 - Выборочной стандартной ошибки
 - Выборочной медианы
5. Для построенных выборок постройте диаграмму рассеяния (https://ru.wikipedia.org/wiki/Диаграмма_рассеяния)
6. Для построенных выборок разделите ось абсцисс на 10 равных отрезков, рассчитайте количество значений в каждом из них и постройте гистограмму (<https://ru.wikipedia.org/wiki/Гистограмма>)
7. Для построенных выборок постройте выборочную функцию распределения и теоретическую функцию распределения. Сравните теоретическое и выборочное распределения

1.2. Центральная предельная теорема (10 баллов)

1. Создайте 100 выборок объемом $n=20$ из равномерного на $[-a_2, a_3]$ распределения и вычислите соответствующие выборочные средние по каждой выборке.
2. Постройте гистограмму стандартизированного выборочного среднего из предыдущего пункта.
3. Сделайте выводы о типе полученного распределения.

1.3. Корреляционные связи статистических рядов (10 баллов)

1. Выберите три произвольных ряда из базы данных цен мировых товарных рынков по адресу <http://pubdocs.worldbank.org/en/561011486076393416/CMO-Historical-Data-Monthly.xlsx>
2. Рассчитайте корреляции по каждой паре из выбранных рядов и сделайте выводы о наличии зависимостей между показателями.

Задание 2. Доверительные интервалы

2.1. Генерация выборок (10 баллов)

1. Сгенерируйте 3 выборки из нормального распределения с параметрами $\mu = a_1$ и $\sigma^2 = a_3$ объемом 100, 300 и 1000 наблюдений
2. Сгенерируйте 3 выборки из экспоненциального распределения с параметром $\lambda = 1/b_2$ объемом 100, 300 и 1000 наблюдений
3. Сгенерируйте 3 выборки из равномерного на $[c_3, c_2 + c_3]$ распределения объемом 100, 300 и 1000 наблюдений

2.2. Доверительные интервалы для одного параметра (10 баллов)

1. Для каждой выборки из пункта 2.1, предполагая ее нормальность, рассчитайте 90% и 95% доверительные интервалы для математического ожидания, предполагая дисперсию известной.
2. Для каждой выборки из пункта 2.1, предполагая ее нормальность, рассчитайте 90% и 95% доверительные интервалы для математического ожидания, предполагая дисперсию неизвестной.
3. Для каждой выборки из пункта 2.1, предполагая ее нормальность, рассчитайте 90% и 95% доверительные интервалы для дисперсии и стандартного отклонения, предполагая математическое ожидание известным.
4. Для каждой выборки из пункта 2.1, предполагая ее нормальность, рассчитайте 90% и 95% доверительные интервалы для дисперсии и стандартного отклонения, предполагая математическое ожидание неизвестным.
5. Сделайте выводы о том, как соотносятся между собой доверительные интервалы в пунктах 1-4.

2.3. Доверительные интервалы для разности параметров (10 баллов)

1. Для каждой пары выборок из каждого распределения из пункта 2.1, предполагая их нормальность, постройте 90% и 95% доверительные интервалы для разностей математических ожиданий, предполагая дисперсии известными.
2. Для каждой пары выборок из каждого распределения из пункта 2.1, предполагая их нормальность, постройте 90% и 95% доверительные интервалы для разностей математических ожиданий, предполагая дисперсии неизвестными, но равными.
3. Преобразуйте 3 выборки из равномерного распределения (пункт 2.1.3) в 3 выборки из биномиального распределения с помощью следующего правила: если элемент выборки меньше $c_2 + \frac{c_1}{c_1 + c_2} c_3$, то ставим 0, если больше или равен $c_2 + \frac{c_1}{c_1 + c_2} c_3$, то ставим 1. Для всех трех пар выборок постройте 90% и 95% доверительные интервалы для всех возможных разностей долей.
4. Сделайте выводы о том, как соотносятся между собой доверительные интервалы в пунктах 1-3.

Задание 3. Проверка гипотез

3.1. Использование симулированных данных (15 баллов)

1. Создайте 100 выборок объемом $n=100$ из нормального распределения с параметрами $\mu = a_2$ и $\sigma^2 = a_1$.
2. Для каждой выборки на 1%, 5% 10% уровне значимости проверьте гипотезу о том, что математическое ожидание равно истинному значению (при известной дисперсии). Рассчитайте долю верных ответов.
3. Для каждой выборки на 1%, 5% 10% уровне значимости проверьте гипотезу о том, что математическое ожидание равно истинному значению (при неизвестной дисперсии). Рассчитайте долю верных ответов.

3.2. Использование реальных данных (15 баллов)

1. Выберите три произвольных ряда из базы данных цен мировых товарных рынков по адресу <http://pubdocs.worldbank.org/en/561011486076393416/CMO-Historical-Data-Monthly.xlsx>
2. На уровне значимости 5% для каждого показателя проверьте гипотезы о том, что его математическое ожидание равно последнему доступному значению (значению за апрель 2023 года или за май 2023 года после его появления в начале июня) на выборках в 36, 72 и 108 последних наблюдений.