# **Индивидуальная работа по курсу**

# **«Теория вероятностей и математическая статистика»**

Целью работы является продемонстрировать способность мыслить и анализировать статистические показатели. Обязательной частью является анализ полученных данных, рассуждения, выводы.

Структура работы:

1. Титульный лист.
2. Введение. Краткие теоретические сведения (определения, теоремы, формулы, общие постановки задач).
3. Расчетная часть: обработка данных, получение статистических показателей и их анализ.
4. Заключение.
5. Список использованных источников.

Оригинальная тема вносится в общий список.

Ответ представляется в формате pdf и дополнительно – расчетные файлы (\*.xslx – при работе в Excel; Gretl script file - при работе в Gretl; Python file, \*.ipynb – при работе в Python.

Максимальный балл – 5.

Критерии оценивания: Оценивается верность вычислений, правильность аналитических заключений и логических рассуждений, общая логика работы, а также способность ясно, грамотно и точно излагать свои мысли в письменном виде. У любого, читающего работу, должно сложиться целостное понятие о рассматриваемом процессе. Оценка снижается за вычислительные и логические ошибки, а также за несвязный, неграмотный, нечитаемый текст.

Оценка хорошо и отлично ставится при условии успешной защиты индивидуальной работы.

**Копии работ (плагиат) оцениваются в 0 баллов для автора и для копирующего.**

**Задачи:**

1. Собрать данные из открытых источников (rosstat, Сбер, сайты министерств и ведомств и т.д., указать источник и дату обращения): 1 объясняемая переменная (Y) и 1 объясняющая переменная (X), выборка пространственная. Количество наблюдений n>=30. Обосновать выбор объясняющей переменной.
2. Определить числовые характеристики статистического распределения Х и Y, описать смысл полученных значений.
3. Найти точечные оценки основных параметров статистических распределений Х и Y.
4. Определить коэффициент вариации и сделать выводы об однородности выборки признака Y.
5. При необходимости провести анализ на статистические выбросы и удалить их (диаграмма «Ящик с усами», статистические критерии).
6. Провести проверку статистической гипотезы о нормальности распределения признака Y в генеральной совокупности.
7. Построить 95% интервальную оценку для среднего признака Y в предположении нормальности распределения признака в генеральной совокупности.
8. Разделить выборку признака Y на две части и провести проверку статистических гипотез о равенстве дисперсий и средних в полученных выборках (двухвыборочные тесты).
9. Построить поле корреляции (X,Y). Добавить на поле корреляции линию тренда, уравнение регрессии и коэффициент детерминации и сформулировать гипотезу о форме взаимосвязи между признаками.
10. Определить коэффициент корреляции rxy, его статистическую значимость, и на этой основе уточнить гипотезу о форме взаимосвязи между признаками.
11. Найти оценки параметров b ̂\_0 и b ̂\_1 уравнения парной линейной регрессии и проверить их значимость при минимально возможном уровне значимости α. Сформулировать содержательную интерпретацию оценок параметров регрессии b ̂\_0 и b ̂\_1 (если это возможно).
12. Проверить значимость модели в целом с помощью F-теста при уровне значимости α.
13. Найти регрессионную, остаточную и общую суммы квадратов (RSS, ESS, TSS) и коэффициент детерминации R^2, сформулировать прикладной смысл коэффициента детерминации.
14. Построить график исходных значений объясняемой переменной и полученных модельных и эмпирическое уравнение регрессии.
15. Построить доверительные интервалы с надежностью γ для истинных значений параметров b\_0 и b\_1. Описать смысл доверительных интервалов с прикладной точки зрения.
16. Проверить остатки построенной модели на гомоскедастичность:

* построить график остатков и сделать его визуальный анализ;
* проанализировать робастные стандартные ошибки;
* провести тесты на гетероскедастичность.

При необходимости выполнить коррекцию гетероскедастичности и построить новую итоговую модель.

1. Для итоговой модели построить график исходных значений объясняемой переменной и полученных модельных и доверительные интервалы с максимально возможной надежностью γ для истинных значений параметров b\_0 и b\_1. Найти точечный и интервальный прогнозы индивидуального значения Y, взяв в качестве прогнозного значения X значение признака, превышающее среднее на 45%. Описать прикладной смысл полученных прогнозов.
2. Сделать выводы по итоговой модели в работе.

Результат работы представить в виде наглядного информативного отчета. Для расчетов использовать ПО: Gretl; Excel; Python.

**Примечание к работе.** Дедлайн 03.06 (но чем раньше, тем лучше). Скриншоты из ПО без подписей и анализа в работе не засчитываются.