

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Статистика в сфере публичной политики (Statistics for Public Policy) наименование дисциплины

Автор: доцент кафедры международного менеджмента, PhD Нейштадт И. В.

Код и наименование направления подготовки, профиля: 41.04.04 Политология

Квалификация (степень) выпускника: Магистр

Форма обучения: Очная

Цель освоения дисциплины:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-3	Углубленное знание общих и специальных методов современной политической науки, уверенное владение навыками применения методологии политической науки к анализу современных политических процессов	ПК-3.1	Знание общих и специальных методов политического анализа.

План курса:

№	Наименование	Содержание
Тема 1	Введение в статистику и исследование	<ul style="list-style-type: none"> • Как можно задавать содержательные вопросы и получать достоверные ответы на них? • Случайные эксперименты и проверка каузальных отношений, наблюдательные (наблюдательные) исследования • Описание работы, исследования и опросники ГА: Введение в STATA (лог файлов, Data Editor, запись параметров)
Тема 2	Предварительный анализ данных – единичная переменная	<ul style="list-style-type: none"> • Виды вариационных рядов, частотное распределение, Графическое изображение (столбчатые диаграммы, Круговые диаграммы, гистограммы) • Способы определения общей тенденции (средняя величина, медиана, мода), дисперсионный анализ (квартили, боксплоты, дисперсия, стандартное (квадратичное отклонение)) • Кривые распределения, нормированное (или стандартное) распределение (таблица Z-преобразования, графики для построения квантилей распределения) STATA: Частотные таблицы, графики и гистограммы. Суммарная статистика
Тема 3	Предварительный анализ данных – связь между переменными	<ul style="list-style-type: none"> • Связь между категориальными независимыми переменными и количественными переменными • Связь между двумя категориальными переменными • Связь между двумя количественными переменными: диаграмма разброса (рассеивания), Корреляция STATA: Таблицы средних величин, кросс-таблицы, точечные диаграммы

№	Наименование	Содержание
Тема 4	Предварительный анализ данных – простая линейная регрессия	<ul style="list-style-type: none"> Простая линейная регрессия (Метод наименьших квадратов, расчет и интерпретация коэффициентов регрессии) Предполагаемые значения и стандартные отклонения, R-квадрат, выбросы (промахи), ассоциированные случайные величины, причинная обусловленность STATA: простая регрессия, простая линейная регрессия
Тема 5	Основы теории вероятности	<ul style="list-style-type: none"> Основные понятия и определения, сложное событие и его вероятность, вероятность совместных событий, вероятность несовместных событий Статистическая (не) зависимость случайных величин, таблицы случайных чисел Случайные величины (дискретные и постоянные), средняя величина и показатели вариации, распределения вероятностей STATA: сопряженные таблицы (кросс-таблицы)
Тема 6	Дисперсия выборки	<ul style="list-style-type: none"> Пример распределения величин (расчет вероятности выборок с биномиальным распределением), пример распределения для соотношений, нормальная аппроксимация биномиального распределения Ряды распределения в статистике. Центральная предельная теорема.
Тема 7	Как делается статистическое заключение	<ul style="list-style-type: none"> t-критерий Стьюдента, Доверительный интервал Проверка гипотезы (нулевая гипотеза и альтернативная гипотеза, одновыборочный и двухвыборочный тесты, Р-значение) STATA: Одновыборочный критерий Стьюдента
Тема 8	Статистическое сравнение двух выборочных средних значений, совместное распределение двух переменных	<ul style="list-style-type: none"> Сравнение двух средних значений (независимые выборки, объединенная дисперсия), доверительный интервал относительных показателей, проверка гипотезы Предположение о взаимосвязи единичных явлений (Доверительные интервалы, проверка гипотезы) Предположение о различии между двумя показателями (CI, проверка гипотезы) STATA: Критерий Стьюдента, одновыборочный критерий Стьюдента, парный выборочный t-тест для средних величин.
Тема 9	Таблицы сопряженности & простая линейная регрессия	<ul style="list-style-type: none"> Таблицы сопряженности (краткий обзор), тест хи-квадрат для определения существенной корреляции между двумя категориальными переменными. Простая линейная регрессия, достоверность совпадения, Условия Гаусса-Маркова Определение коэффициента наклона (стандартная ошибка, доверительные интервалы, р-значение) STATA: Кросс-таблицы, Хи-квадрат, оценка простой регрессии, оценка статистической значимости, расчет предполагаемых значений и остаточной погрешности, достоверность
Тема 10	Множественная линейная регрессия I	<ul style="list-style-type: none"> Случайные наблюдения и эксперименты, определяющие факторы Поиск подходящей модели, проверка предположений о регрессии, проверка модели. F-статистика, коэффициент детерминации (R-квадрат) STATA: оценка множественной линейной регрессии, анализ регрессионных остатков
Тема 11	Множественная линейная регрессия II	<ul style="list-style-type: none"> Структурные (фиктивные) переменные (одинаковый коэффициент наклона, различные пересечения) Зависимые переменные (различные коэффициенты наклона, различные пересечения) Категориальные переменные STATA: Работа с фиктивными переменными и зависимыми переменными.

Формы текущего контроля и промежуточной аттестации:
Примеры задач в классе:

Примеры задач, решаемых на занятиях:

Задача 1. (10 баллов.) Покерная комбинация

Самой часто встречающейся комбинацией в картах является комбинация из одной пары идентичных карт, за ней следуют комбинация из двух и трех пар. Комбинация из двух пар: две одинаковых карты одной масти, две одинаковых карты другой масти, оставшаяся карта имеет третью масть, например $\{2♥, 2♠, 5♥, 5♣, K♦\}$

Тройка карт одинакового достоинства: Три одинаковые карты, остальные две отличаются, например $\{2♥, 2♠, 2♣, 5♣, K♦\}$

Рассчитайте вероятность каждого вида комбинаций? Какая из них наиболее вероятная?

Задача 2. (10 баллов.) Нестандартный (нетранзитивный) игральный кубик.

На занятии мы работали с нестандартным (нетранзитивным) игральным кубиком:

Красный: 3 3 3 3 3 6; Зеленый: 1 4 4 4 4 4; Белый: 2 2 2 5 5 5.

Закончите делать Ваш нестандартный кубик.

(а) На занятии Вы выяснили, что белый сильнее белого. Выясните, насколько истинно предположение о том, что белый сильнее зеленого и что зеленый сильнее красного.

Укажите все варианты (по порядку, от лучшего к худшему), которые возможны при бросании кубика от лучшего к худшему? (Подсказка: мы имеем дело с «отсутствием транзитивности»)

(b) Предположим, что Вы бросаете два белых кубика против двух красных кубиков. Какова вероятность того, что сумма значений белого кубика будет больше суммы значений красного кубика?

Задача 3. (20 баллов) Дни рождения: расчет и моделирование.

Пронумеруем все дни года (невисокосного) от 1 до 365. Допустим, что дни рождения могут с одинаковой степенью вероятности приходиться на любой день года. Рассмотрим группу из n человек (Вы в нее не входите). An element of the sample space Ω will be a последовательность из n дней рождения (один день для каждого).

(а) Определите функцию распределения вероятности P для Ω .

(b) Рассмотрим следующее:

A: “у кого-то из участников группы день рождения совпадает с Вашим”

B: “у кого-то двоих в группе дни рождения совпадают”

C: “у троих из группы день рождения в один и тот же день”

Подробно опишите подмножество Ω в каждом случае.

(c) Подберите точную формулу для $P(A)$. Какое самое маленькое n , чтобы $P(A) > .5$?

(d) Объясните, почему n больше 365, не производя расчетов (дайте краткий ответ, полагаясь на логику и/или интуицию)

(e) Осуществите имитационное моделирование для оценки наименьшего значения n , для которого $P(B) > .9$. Для составления этих моделей, количество анализов 10000.

Для этого значения n , осуществите моделирование несколько раз, чтобы убедиться в том, что результат каждый раз будет одинаковым.

Осуществив 10000 выполненных анализов, Вы убедились, что в оценке $P(B)$ отклонение значения очень небольшое. Прodelайте это еще раз, осуществив 30 анализов, и убедитесь в том, что вероятность ошибки в этом случае больше.

(f) Подберите точную формулу для $P(B)$.

(g) Воспользуйтесь имитационным моделированием, чтобы оценить наименьшее значение n , для которого $P(C) > .5$. Снова осуществите 10000 анализов. Вы убедитесь в том, что два значения n одинаково достоверны. Вы можете использовать одно из них в ответе.

Типовые оценочные средства

Письменный экзамен: примеры задач

1. Существует три варианта расположения букв слова «DAD» (папа), а именно DAD, ADD, и DDA. Сколько вариантов различного расположения букв слова «PROBABILITY» (вероятность)?

2. На занятии бальными танцами присутствуют шесть мужчин и семь женщин. Если выбрать четырех женщин и четырех мужчин и поставить их в пары, сколько пар можно составить?

3. Обозначим два события буквами А и В. Предположим, что вероятность того, что ни А ни В не произойдет, равняется $2/3$. Какова вероятность того, что произойдет или одно или оба события?

4. Обозначим два события буквами С и D, где $P(C) = 0.25$, $P(D) = 0.45$, и $P(C \cap D) = 0.1$.

Каково значение $P(C \cap D)$?

5. Представьте, что Вы сдаете тест с вариантами ответов. При ответе вероятность того, что Вы дадите правильный ответ, равна p . Если Вы не знаете ответ, Вы выбираете произвольный вариант. Какова вероятность того, что если бы Вы знали ответ, Вы бы ответили правильно?

6. Бросили две игральных кубика

A = 'сумма очков на двух кубиках равняется 3'

B = 'сумма очков на двух кубиках равняется 7'

C = 'по крайней мере на одном из кубиков выпало 1 очко'

(a) Что означает $P(A|C)$?

(b) Что означает $P(B|C)$?

(c) А и С независимые величины? Что можно сказать о В и С?

7. Во время экзамена студенту необходимо выбрать правильный ответ из четырех вариантов. Студент хорошо подготовился. Вероятность того, что он будет знать ответ на вопрос, равна 0.5, вероятность того, что он сможет исключить один вариант ответа -0.25, в остальных случаях все четыре ответа кажутся одинаково вероятными.

Если он знает ответ, он выберет правильный вариант. Если нет, придется сделать выбор из 3 или 4 представленных. Вы учитель и Вам предстоит оценить знания студента. Если студент отвечает на вопрос правильно, какова вероятность осознанного (не случайного) выбора?

8. Предположим, что $P(A) = 0.4$, $P(B) = 0.3$ and $P((A \cup B)C) = 0.42$. Есть ли зависимость между А and В?

9. Предположим, что X принимает значение между 0 и 1 и имеет интегральную функцию $2x$.

Вычислите $\text{Var}(X)$ и $\text{Var}(X^2)$.

10. Рассчитайте вероятность и отклонение бернуллиевской случайной величины (p)

Основная литература:

1. Rohatgi, Vijay K., and A. K. Ehsanes Saleh. An Introduction to Probability and Statistics, John Wiley & Sons, Incorporated, 2015. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/ranepa-ebooks/detail.action?docID=4038785>.
2. Hanneman, Robert A., et al. Basic Statistics for Social Research, John Wiley & Sons, Incorporated, 2012. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/ranepa-ebooks/detail.action?docID=861634>.