

ОП «Политология», 2022-23**Введение в ТВиМС****Тренировочные задания по блоку «Статистика»****Не является типовым вариантом экзамена! (4 модуль)***А. А. Макаров, А. А. Тамбовцева***Задача 1.** Дана выборка: 4, 18, 9, 10, -1, 9, 25, 18, 9

- (a) Запишите вариационный ряд и ранги наблюдений.
- (b) Найдите медиану этой выборки.
- (c) Найдите верхний и нижний квартили этой выборки.
- (d) Проверьте, есть ли в выборке нетипичные значения (выбросы). Если нет, напишите, что их нет, если есть – перечислите.
- (e) Постройте для данной выборки «ящик с усами».
- (f) Найдите выборочное среднее и выборочное стандартное отклонение.

Задача 2. Представьте, что вам необходимо построить доверительный интервал для средней стоимости квартиры (в млн рублей) в Москве площадью 30-40 квадратных метров. Известно, что доверительный интервал будет строиться на основе стоимости 23 квартир, средняя стоимость квартир по такой выборке равна 20 млн рублей, стандартное отклонение равно 4 млн рублей. Постройте доверительный интервал для средней стоимости квартиры, приняв уровень доверия равным: а) 90%; б) 95%; в) 99%.

Задача 3. Исследователь оценивает долю жителей страны А, которые поддерживают новый закон о налогообложении. Он опросил 1600 человек и выяснил, что 1000 человек из них поддерживают новый закон. Постройте 92%-ный доверительный интервал для доли жителей, поддерживающих новый закон. Что произойдет с длиной доверительного интервала, если увеличить объем выборки в 4 раза (при прочих равных условиях)?

Задача 4. Группа политологов проводит исследование, посвященное политическим предпочтениям молодежи. Выяснилось, что 62% респондентов придерживаются либеральных взглядов (всего было опрошено 1000 человек). Вам необходимо проверить гипотезу о равенстве доли приверженцев либеральных взглядов 0.55.

- (a) Сформулируйте нулевую гипотезу. Сформулируйте альтернативную гипотезу, считая, что она двусторонняя.
- (b) Какое распределение имеет статистика критерия, используемого для проверки нулевой гипотезы?
- (c) Вычислите наблюдаемое значение статистики критерия и p-value. Проверьте гипотезу на 5% уровне значимости. Сделайте статистический и содержательный вывод.
- (d) Изменится ли вывод относительно нулевой гипотезы, если мы примем уровень значимости равный 10%? Уровень значимости равный 1%?

Задача 5. Даны два показателя: явка на выборы (в процентах) и доля трудоспособного населения (в процентах), посчитанные по 28 регионам России.

- (a) Представьте, что вам необходимо проверить статистически, есть ли связь между этими двумя показателями. Вы посчитали коэффициент корреляции Пирсона. Сформулируйте нулевую гипотезу, которую необходимо проверить, чтобы выяснить, значим ли коэффициент корреляции. Какое распределение имеет статистика критерия, используемого для проверки гипотезы? С какими параметрами?
- (b) Какое значение должно быть у p-value, чтобы коэффициент корреляции считался значимым на 5% уровне значимости?

Задача 6. Два эксперта оценивают эссе абитуриентов по английскому языку. Оценки выставляются в 10-балльной шкале. Перед вами оценки, поставленные шести абитуриентам:

Эксперт 1: 10, 8, 6, 4, 3, 9

Эксперт 2: 8, 9, 3, 5, 4, 10

- (a) Рассчитайте коэффициент корреляции Спирмена между оценками, поставленными двумя экспертами.
- (b) Проверьте гипотезу о независимости признаков между показателями на 1% уровне значимости. Сделайте статистический и содержательный вывод. Можно ли считать, что эксперты ставят оценки согласованно?

Задача 7. Дана таблица сопряженности двух признаков: *пол респондента* и его ответ на вопрос: «*Любите ли вы зелёный чай?*».

	Да	Нет
Женский	40	10
Мужской	34	8

- (a) Вам необходимо проверить, есть ли связь между этими признаками. Сформулируйте нулевую и альтернативную гипотезу. Какое распределение имеет статистика критерия, используемого для проверки нулевой гипотезы?
- (b) Посчитайте наблюдаемое значение статистики и p-value. Какой вывод относительно нулевой гипотезы можно сделать на 5% уровне значимости?

Задача 8. Известно, что коэффициент корреляции Пирсона между индексом политических свобод (принимает значения от 1 до 100) и ВВП на душу населения равен 0.76. Проверьте на 1% уровне значимости гипотезу о равенстве истинного значения коэффициента корреляции нулю (p-value можно оценить качественно, не досчитывая до точного значения), если известно, что $n = 9$.

Задача 9. Известно, что оценки студентов за тест по курсу «Категории политической науки» имеют нормальное распределение со средним значением 6 и дисперсией 4. Из всех студентов независимо выбирают 100 выборок по 25 человек в каждой.

- (a) Какое распределение будет иметь набор средних значений, посчитанных по этой выборке? С какими параметрами?
- (b) Найдите вероятность того, что средняя оценка за тест, полученная по выборке в 25 человек, превысит значение 4.

Задача 10. В Python был посчитан коэффициент корреляции Пирсона между явкой (в процентах) и процентом голосов за кандидата А. Выдача с результатом выглядит следующим образом:

```
coef=0.6968 pvalue=0.0251
```

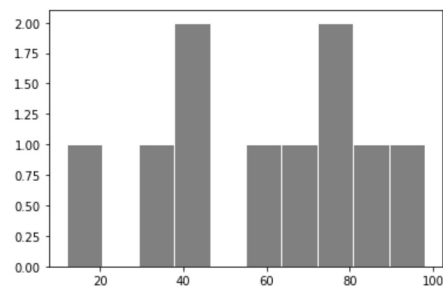
- (a) Проинтерпретируйте полученные результаты: укажите направление и силу связи.
- (b) Можно ли считать на 1% уровне значимости, что связь между показателями существует?
- (c) Как изменится значение p-value, если мы сделаем альтернативную гипотезу односторонней? Повлияет ли это на вывод в предыдущем пункте?

Задача 11. В Python была проверена гипотеза о равенстве среднего возраста сторонника кандидата Незнама Незнамовича Незнайкина 50 годам. Выдача с результатом выглядит следующим образом:

```
Ttest_1sampResult(statistic=2.093223452953117, pvalue=0.04518732048925306)
```

- (a) Сформулируйте нулевую и одностороннюю альтернативную гипотезу (направление выберите исходя из данных).
- (b) На основе выдачи сделайте вывод о нулевой гипотезе, приняв уровень значимости равным 10%.
- (c) Как изменится значение p-value, если мы сделаем альтернативную гипотезу двусторонней? Повлияет ли это на вывод в предыдущем пункте?

Задача 12. Для некоторой выборки в Python была построена гистограмма:



- (a) Как изменится эта гистограмма, если мы добавим в выборку значения 110 и 60?
- (b) Можно ли считать распределение этой выборки нормальным?