ОП «Политология», 2017-18
Математика и статистика, часть 2
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. Проинтерпретируйте полученное занчение.
(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) Постройте гистограмму для данной выборки, выбрав стартовым значением минимальное значение в выборке и интервал группировки равный 4.
(g) Найдите выборочное среднее и выборочное стандартное отклонение (несмещенную оценку).

Задача 2. Дана выдача R, в которой представлено описание переменной *число лошадиных сил автомобиля*:

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
52.0    96.5   123.0   146.7   180.0   335.0
```

Поясните, что представляет собой каждое значение в выдаче.

Задача 3. Представьте, что вам необходимо построить доверительный интервал для средней стоимости квартиры (в рублях) в Центральном округе Москвы. Известно, что доверительный интервал будет строиться на основе стоимости 26 квартир Центрального округа.

Какое распределение имеет статистика, на значение которой будет умножаться стандартная ошибка выборки при расчете доверительного интервала? Найдите значение этой статистики, приняв уровень доверия равным: а) 90%; б) 95%; в) 99%; г) 99.5%.

Задача 4. 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 5. Группа политологов проводит исследование, посвященное политическим предпочтениям молодежи. Выяснилось, что 62% респондентов придерживаются либеральных взглядов. 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) Наблюдаемое значение статистики критерия равно 2.19. Проверьте гипотезу на 5% уровне значимости, построив критическую область для значений критерия. Сделайте статистический и содержательный вывод.
(d) Рассчитайте p-value. Что означает это значение? Есть ли основания отвергнуть нулевую гипотезу?
(e) Изменится ли вывод относительно нулевой гипотезы, если мы примем уровень значимости равный 10%? Уровень значимости равный 1%?

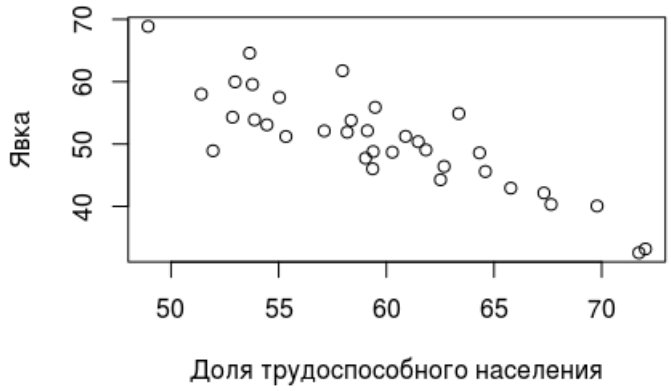
Задача 6. Проводится исследование, посвященное уровню жизни в регионах Российской Федерации. Разработана методика оценки, сконструирован интегральный индекс уровня жизни. Проверяется гипотеза о равенстве индекса уровня жизни в регионах европейской части России и азиатской части России.

(a) Сформулируйте нулевую гипотезу и правостороннюю альтернативную гипотезу.
(b) Какое распределение имеет статистика критерия, используемого для проверки нулевой гипотезы? Постройте критическую область для статистики критерия, считая, что нулевая гипотеза проверяется на 10% уровне значимости.
(c) Выдача R по результатам проверки нулевой гипотезы выглядит так:

```
data:  e and a
t = 7.9266, df = 138.93, p-value = 6.613e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  6.47209 10.77389
sample estimates:
mean of x mean of y
 74.12198  65.49899
```

Какой вывод, касающийся нулевой гипотезы, можно сделать, если мы проверяем ее на 10% уровне значимости? Сделайте содержательный вывод.

Задача 7. Дана диаграмма рассеяния явки на выборы (в процентах) и доли трудоспособного населения (в процентах):



(a) Какой вывод о связи двух показателей можно сделать? Можно ли по этому графику сделать вывод, что процент трудоспособного населения оказывает отрицательное влияние на явку?
(b) 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?
(c) Какое значение должно быть у p-value, чтобы коэффициент корреляции считался значимым на 5% уровне значимости?

Задача 8. 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 равенстве коэффициента корреляции нулю. Сделайте статистический и содержательный вывод. Можно ли считать, что эксперты ставят оценки согласованно?

Задача 9. Дана таблица сопряженности двух признаков: *пол респондента* и его ответ на вопрос: «*Любите ли вы горький шоколад?*».

	Да	Нет
Женский	32	14
Мужской	15	25

(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) Известно, что p-value равно 0.006. Какой вывод относительно нулевой гипотезы можно сделать?

Задача 10. Известно, что коэффицент корреляции Пирсона между индексом политических свобод (принимает значения от 1 до 100) и ВВП на душу населения равен 0.76. Проверьте на 5% уровне значимости гипотезу о равенстве истинного значения коэффициента корреляции нулю, построив критическую область для значений статистики используемого критерия.

Задача 11. Найдите:

(a) $P(Z < 1.12)$, $P(Z > 2.32)$, $P(Z < 4.82)$, $P(Z < -0.89)$.

(b) $P(T < 0)$, $P(T > 0)$.

где Z – стандартная нормальная величина, T – величина, имеющая распределение Стьюдента с $df = 12$.

Задача 12. 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 100 человек в каждой. Какое распределение будет иметь набор средних значений, посчитанных по этой выборке? С какими параметрами?

Задача 13. Сотрудники социологического агенства «НЕ (В)ОПРОС» по результатам опроса выяснили, что 87% респондентов верят в то, что если дорогу перебежит черная кошка – это к неудаче. Найдите вероятность того, что в выборке из 1200 респондентов будет от 600 до 1000 человек, которые верят в эту примету.