

Пакет стандартных задач

Задача 2

Вариант 1

Имеются несгруппированные данные выборочного наблюдения (30% случайная выборка) использования служебных сотовых телефонов в компании «Альфа» (минут разговоров в месяц):

Phone Use	Phone Use
346	505
442	515
442	429
480	358
345	466
466	549
461	470
490	471
489	372
505	437
515	516

1. Протестируйте выборочную совокупность на наличие выбросов, рассчитав межквартильное расстояние (IQR).
2. Рассчитайте структурные средние для приведенной выборки.
3. Рассчитайте несмещенную дисперсию выборки, пользуясь формулой

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{(\sum f_i) - 1} . \text{ Объясните, почему для выборочной совокупности необходимо}$$

рассчитывать несмещенную дисперсию?

4. Рассчитайте ошибку дисперсии, если известно, что $\sigma^2 = 3500$.
5. Можно ли считать с 95% вероятностью, что генеральная средняя будет равна 450 минутам?

Вариант 2

Имеются результаты первичного анализа данных сплошного опроса студентов бакалавриата 4 года обучения:

Распределение студентов 4 курса по числу часов, отработанных ими в свободное от учебы время (часов в неделю)

Время, затраченное студентами на подработку (часов в неделю)	Число студентов
0-9	44
10-19	24
20-29	59
30-39	28
40-49	5

1. Дополните результаты анализа графиком, наиболее подходящим для данных в таблице. Дайте несколько идей для дальнейшего анализа данных.
2. Рассчитайте среднее, модальное и медианное время, отработанное студентами, а также коэффициент вариации.
3. Применяя эмпирическое правило, сделайте вывод о том, можно ли считать распределение студентов приближенно нормальным.
4. Определите, какая доля студентов попадет в интервал $\pm 2\sigma$?

Вариант 3

В таблице приведены результаты социологического опроса, проведенного в сельскохозяйственном регионе. Целью опроса было изучение взаимосвязи между удовлетворенностью местоположением своего домовладения (частного дома, фермы) и сводным индексом качества жизни (интегральная характеристика, включающая в себя загрязненность водоемов, сервис и доступность культурных и торговых объектов, уровень шума и т.п., всего 20 показателей).

Сопряженность между качеством жизни и удовлетворенностью местоположением своего домовладения для жителей N-ского региона

Удовлетворенность	Качество жизни			Итого
	Низкое	Среднее	Высокое	
Низкая	21	15	6	42
Средняя	12	25	21	58
Высокая	8	17	32	57
Итого	41	57	59	157

1. Рассчитайте фактическое значение критерия Хи-квадрат и протестируйте гипотезу о связи между признаками. Сделайте вывод.
2. Известно, что для таблиц размерностью 2×2 , гипотеза о связи тестируется по скорректированному критерию с поправкой Йетса, рассчитываемому по формуле:

$$\chi_c^2 = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}, \text{ где } f_o - \text{наблюдаемая частота, а } f_e - \text{ожидаемая}$$

(теоретическая) частота. Перегруппируйте данные, используя 2 категории для качества жизни («высокое» и «среднее и низкое») и удовлетворенности («высокая» и «средняя и низкая»). Рассчитайте скорректированный хи-квадрат, протестируйте новую гипотезу о связи. Сделайте выводы.

Вариант 4

Имеется информация о выпуске продукции (работ, услуг), полученной на основе 10% выборочного наблюдения по предприятиям области:

Группы предприятий по объему продукции, тыс. руб.	Число предприятий (f)
1	2
До 100	28
100-200	52
200-300	164
300-400	108
400-500	36
500 и >	12
итого	400

Определить:

1) по предприятиям, включенным в выборку: а) средний размер произведенной продукции на одно предприятие; б) дисперсию объема производства; в) долю предприятий с объемом производства продукции более 400 тыс. руб.;

2) в целом по области с вероятностью 0,954 пределы, в которых можно ожидать: а) средний объем производства продукции на одно предприятие; б) долю предприятий с объемом производства продукции более 400 тыс. руб.;

3) общий объем выпуска продукции по области.

Вариант 5

Компания «Зебра Принт» анализирует сроки оплаты клиентами, имеющими кредитные лимиты, выставленных счетов. Бухгалтер предполагает, что сроки оплаты всех счетов нормально распределены и имеют $\mu = 47$ дней и $\sigma = 11$ дней. Для оценки аудитор выбрал в сентябре случайным образом 10 счетов, сроки оплаты по которым составили (в днях):

55 45 45 42 65

58 35 36 34 60

1. Используйте данные выборки для получения точечной оценки средней.
2. Постройте 95% интервал для среднего числа дней, которое требуется клиентам для оплаты счетов. Сделайте выводы.
3. Протестируйте гипотезу $H_0 : \mu = 47$ при уровне значимости $\alpha = 0.05$.