

# Содержание

	Стр.
<b>1. 1</b> .....	<b>2</b>
<b>2. 2</b> .....	<b>2</b>
<b>3. 3</b> .....	<b>2</b>
<b>4. 4</b> .....	<b>2</b>
<b>5. 5</b> .....	<b>3</b>
<b>6. 6</b> .....	<b>3</b>
<b>7. 7</b> .....	<b>3</b>
<b>8. 8</b> .....	<b>3</b>
<b>9. 9</b> .....	<b>4</b>
<b>10.10</b> .....	<b>4</b>
<b>11.11</b> .....	<b>4</b>
<b>12.12</b> .....	<b>5</b>
<b>13.13</b> .....	<b>5</b>
<b>14.15</b> .....	<b>5</b>
<b>15.16</b> .....	<b>5</b>
<b>16.17</b> .....	<b>5</b>
<b>17.18</b> .....	<b>6</b>
<b>18.19</b> .....	<b>6</b>

## 1 1

### Формулировка

Формы представления данных и цели анализа данных.

### Ответ

По форме представления данных можно выделить выборки, функции, семантические данные (тексты, изображения). В качестве основных целей анализа данных можно отметить анализ глобальных событий и явлений, компактное представление данных, прогнозирование.

## 2 2

### Формулировка

Характеристики положения данных.

### Ответ

1) Выборочное среднее, 2)  $\text{med } x = x$  ?  $x = x[k+1] \parallel x = x[k]$ , в случае нечётности и чётности соответственно (выборочная медиана), 3)  $Z_R$  = полусумма экстремальных значений, 4) Полусумма выборочных квартилей  $Z_Q = \frac{z_{1/4} + z_{3/4}}{2}$ , 5)  $Z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x(i)$  - усечённое среднее, 6) Среднее геометрическое  $Z_G = x_1 * \dots * x_n^{1/n}$ , 7) Среднее гармоническое  $Z_M = \frac{1}{1/n * \sum_{i=1}^n 1/z_i}$ , 8) Среднее по Колмогорову  $z_k = g^{-1}(\frac{1}{n} \sum_{i=1}^n g(x_i))$ .  
Вариационный ряд данных - неубывающая выборка.

## 3 3

### Формулировка

Характеристики рассеивания данных.

### Ответ

1) Среднеквадратичное отклонение и выборочная дисперсия, 2) среднее абсолютное отклонение от медианы  $d = 1/n \sum_{i=1}^n |x_i - \text{med } x|$ , 3)  $R = X_n - X_1$  - размах выборки, 4) интервальная ширина, 5) Медианное абсолютное отклонение

## 4 4

### Формулировка

Оптимизационный подход к построению x-к положения и рассеивания данных.

**Ответ**

**5 5**

**Формулировка**

Характеристики взаимосвязи данных.

**Ответ**

Коэффициент корреляции Пирсона (линейная зависимость), ранговый коэффициент Спирмена - мера монотонной зависимости (не линейной), квадратный (знаковый) коэффициент корреляции.

**6 6**

**Формулировка**

Характеристики экстремальных значений данных.

**Ответ**

Нужно для выявления выбросов.

1)  $|x_i - medx| > K * MAD$   $x \rightarrow IQR$

2)  $x_i$  - выброс, если  $|x_i - LQ - 3/2IQR| < \min(x_n, UQ + 3/2IQR)$ .

**7 7**

**Формулировка**

Графическое представление данных — боксплот Тьюки.

**Ответ**

$IQR = UQ - LQ$ .  $x_i$  - выброс, если  $|x_i - LQ - 3/2IQR| < \min(x_n, UQ + 3/2IQR)$ .

**8 8**

**Формулировка**

Характеристики распределений данных: «ядерные» оценки плотности.

**Ответ**

Эмпирическая функция распределения, дельта функция.

Ядро - функция  $K(u)$ , если : 1.  $K(u) \geq 0$ , 2.  $K(-u) = K(u)$ , 3.  $\int_{-\infty}^{\infty} K(u) du = 1$ .

Если функция обладает первым свойством, то результатом ядерной оценки плотности действительно будет плотность вероятности. Третье свойство гарантирует, что среднее значение распределения равно среднему использованной выборки.

Нужно для оценки плотности распределения.  $f(x) = \lim_{n \rightarrow \infty, h \rightarrow 0} 1/n * h \sum_{i=1}^n K(\frac{x-x_i}{h})$

## 9 9

### Формулировка

Что такое точечная оценка?

### Ответ

Оценка параметра — соответствующая числовая характеристика, рассчитанная по выборке. Оценки параметров генеральной совокупности делятся на два класса: точечные и интервальные.

Когда оценка определяется одним числом, она называется точечной оценкой. Точечная оценка, как функция от выборки, является случайной величиной и меняется от выборки к выборке при повторном эксперименте.

К точечным оценкам предъявляют требования, которым они должны удовлетворять, чтобы хоть в каком-то смысле быть «доброкачественными». Это несмещённость, эффективность и состоятельность.

Несмещённость - если мат ожидание оценки равно оцениваемому параметру генеральной совокупности. Эффективность - если обладает минимальной дисперсией среди всех несмещенных точечных оценок. Состоятельность - если при увеличении выборки стремится по вероятности к параметру генеральной совокупности. Генеральная совокупность - параметр от которого зависит выборка.

## 10 10

### Формулировка

Что такое статистика?

### Ответ

Статистика — отрасль знаний, наука, в которой излагаются общие вопросы сбора, измерения, мониторинга, анализа массовых статистических (количественных или качественных) данных и их сравнение; изучение количественной стороны массовых общественных явлений в числовой форме.

## 11 11

### Формулировка

Какая оценка называется состоятельной, несмещенной, эффективной, робастной?

### Ответ

Несмещённость - если мат ожидание оценки равно оцениваемому параметру генеральной совокупности. Эффективность - если обладает минимальной дисперсией среди всех несмещенных точечных оценок. Состоятельность - если при увеличении выборки стремится по вероятности к параметру генеральной совокупности. Генеральная совокупность - параметр от которого зависит выборка. Робастная - устойчивая к выбросам

**12 12**

**Формулировка**

Какая из двух оценок считается более эффективной?

**Ответ**

ДЛЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН ВЕРОЯТНОСТЬ ТОГО, ЧТО ТОЧЕЧНАЯ ОЦЕНКА (ширина доверительного интервала равна 0) СОВПАДЕТ С ЛЮБЫМ ЗАДАННЫМ ЗНАЧЕНИЕМ ИЛИ ОЦЕНИВАЕМЫМ ПАРАМЕТРОМ РАВНА 0.

Таким образом, точечная оценка имеет смысл лишь тогда, когда приведена характеристика рассеяния этой оценки (дисперсия). В противном случае она может служить лишь в качестве исходных данных для построения интервальной оценки.

Интервальная оценка лучше.

**13 13**

**Формулировка**

Что такое эффективность, относительная эффективность, асимптотическая эффективность оценки?

**Ответ**

**14 15**

**Формулировка**

Приведите примеры состоятельных оценок м.о. нормального распределения.

**Ответ**

**15 16**

**Формулировка**

Приведите примеры состоятельных оценок м.о. распределения Лапласа.

**Ответ**

**16 17**

**Формулировка**

Приведите примеры состоятельных оценок м.о. равномерного распределения.

**Ответ**

**17 18**

**Формулировка**

Приведите примеры состоятельных оценок центра симметрии распределения Коши.

**Ответ**

**18 19**

**Формулировка**

**Ответ**