Задания к лекции 2. Оценивание параметров.

Вариант 1

Задача 1. Найти эмпирическую функцию по данному расраспределению выборки (x_i – значения, n_i – частоты):

x_i	4	7	8
n_i	5	2	3

Задача 2. Найти точечные оценки математического ожидания, дисперсии для данных `data_p2_var1`.

Задача 3. Вычислить точечные оценки математического ожидания (данные: `data_p3_var1`), исходя из выборки из нормального распределения с параметрами $\mathcal{N}(3,2)$:

$$\sum_i x_i,$$
$$\text{median},$$
$$\max_i x_i,$$
$$\min_i x_i.$$

Какая из перечисленных оценок является наиболее точной (учитывая, что распределение, для которого получена выборка известно).

Задача 4. Сгенерировав 1000 тестовых выборок из нормального распределения $\mathcal{N}(0,1)$ размером 3 убедитесь, что выражение для оценки дисперсии $\hat{\sigma}^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$, в отличие от $\frac{1}{n}\sum_i (x_i - \overline{x})^2$ является более точной (правильнее несмещенной) оценкой σ^2.

Задача 5. Найти выборочную дисперсию по данному расраспределению выборки объема $n = 100$:

x_i	2502	2804	2903	3028
n_i	8	30	60	2

Задача 6. Рассматриваются две оценки математического ожидания: среднее арифметическое значений и медиана. Выполнив имитационное моделирование – создав 500 выборок размером 100 из равномерного распределения на интервале $[0,1]$, дать оценку какая из них более эффективна (т.е. обладает меньшей дисперсией).

Задача 7. Оценкой максимального правдоподобия называется такая оценка, при которой наблюдаемая реализация наиболее вероятна. Найти оценку максимального правдоподобия приживаемости саженцев, если из 100 саженцев прижилось 60.

Задача 8. Построить интервальные оценки дисперсии и математического ожидания для выборки `bigdataset_var1`, полученной из нормального распределения.

Задача 9. Найти минимальный объем выборки, при котором с надежностью 0.975 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta = 0.2$, если известно среднее квадратическое отклонение $\sigma = 1.1$ нормально распределенной генеральной совокупности.

Задача 10. В 360 испытаниях, в каждом из которых вероятность появления события одинакова и неизвестна, событие A появилось 270 раз. Найти доверительный интервал, покрывающий неизвестную вероятность P с надежностью 0.95.

Задача 11. Исследуется распределение безразмерного параметра некоторого биологического объекта (например, отношение длины и ширины листовой пластинки, длины и ширины семени и пр.). Возможны два подхода к вычислению среднего отношения: $m_1 = \dfrac{\sum\limits_i l_i}{\sum\limits_j h_j}$ и $m_2 = \text{median}\left\{\dfrac{l_i}{h_i}\right\}$, $i = \overline{1 \dots n}$. На основе статистического моделирования с выборками из нормального распределения установить какой из этих подходов приводит к более эффективной оценке отношения «длины» (l) и «высоты» (h) объекта.

Задача 12. Создать выборку размером 100, полученные из так называемого «усеченного» нормального распределения с параметрами $\mathcal{N}(1, 1)$ и интервалом определения $[0.5, 1.5]$. Результат представить в виде текстового файла.

Задача 13. Какова вероятность попадания выборочных данных за пределы 2σ-интервала (где σ^2 дисперсия распределения) относительно среднего для равномерного распределения?

Задача 14. Имеются наборы измерений длины и ширины прямоугольного участка, содержащие инструментальные ошибки. Если $\{l_i\}$ и $\{w_i\}$ наборы измерени длины и ширины, то оценку площади можно получить, например, двумя способами: $\dfrac{1}{n}\sum\limits_i l_i \cdot w_i$ и $\text{median}\{l_i\} \cdot \text{median}\{w_j\}$. Провести имитационное моделирование с нормально распределенными величинами с целью установить, какой из подходов наиболее эффективен (оценка площади имеет наименьшую дисперсию).

Задача 15. Выяснить путем имитационного моделирования с выборками из равномерного распределения (на интервале $[-10, 10]$) величину погрешности при оценивании математического ожидания и дисперсии, привносимую вследствие группировки данных. Предположить, что группировка осуществляется по 10 равным интервалам на инервале $[-10, 10]$.

Задания к лекции 2. Оценивание параметров.

Вариант 2

Задача 1. Найти эмпирическую функцию по данному расраспределению выборки (x_i – значения, n_i – частоты):

x_i	2	5	7	8
n_i	1	3	2	4

Задача 2. Найти точечные оценки математического ожидания, дисперсии для данных `data_p2_var2`.

Задача 3. Вычислить точечные оценки математического ожидания (данные: `data_p3_var2`), исходя из выборки из нормального распределения с параметрами $\mathcal{N}(5, 3)$:

$$\sum_i x_i,$$
$$\text{median},$$
$$\max_i x_i,$$
$$\min_i x_i.$$

Какая из перечисленных оценок является наиболее точной (учитывая, что распределение, для которого получена выборка известно).

Задача 4. Сгенерировав 1000 тестовых выборок из нормального распределения $\mathcal{N}(0, 1)$ размером 5 убедитесь, что выражение для оценки дисперсии $\hat{\sigma}^2 = \dfrac{1}{n-1}\sum_i (x_i - \overline{x})^2$, где $\overline{x} = \dfrac{1}{n}\sum_i x_i$ в отличие от $\dfrac{1}{n}\sum_i (x_i - \overline{x})^2$ является более точной (правильнее – несмещенной) оценкой σ^2.

Задача 5. Найти выборочную дисперсию по данному расраспределению выборки объема $n = 100$:

x_i	340	360	375	380
n_i	20	50	18	12

Задача 6. Рассматриваются две оценки математического ожидания: среднее арифметическое значений и медиана. Выполнив имитационное моделирование – создав 400 выборок размером 120 из нормального распределения $\mathcal{N}(1, 1)$, дать оценку какая из них менее эффективна (т.е. обладает большей дисперсией).

Задача 7. Оценкой максимального правдоподобия называется такая оценка, при которой наблюдаемая реализация наиболее вероятна. Найти оценку максимального правдоподобия приживаемости саженцев, если из 100 саженцев прижилось 30.

Задача 8. Построить интервальные оценки дисперсии и математического ожидания для выборки `bigdataset_var2`, полученной из нормального распределения.

Задача 9. Найти минимальный объем выборки, при котором с надежностью 0.975 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta = 0.3$, если известно среднее квадратическое отклонение $\sigma = 1.2$ нормально распределенной генеральной совокупности.

Задача 10. Произведено 300 испытаний по схеме Бернулли. Событие A появилось в 250 испытаниях. Найти доверительный интервал, покрывающий неизвестную вероятность $P(A)$ с надежностью 0.95.

Задача 11. Исследуется распределение безразмерного параметра некоторого биологического объекта (например, отношение длины и ширины листовой пластинки, длины и ширины семени и пр.). Возможны два подхода к вычислению среднего отношения: $m_1 = \dfrac{\sum_i l_i}{\sum_j h_j}$ и $m_2 = \dfrac{1}{n}\sum_{i=1}^{n}\dfrac{l_i}{h_i}$. На основе статистического моделирования с выборками из нормального распределения установить какой из этих подходов приводит к более эффективной оценке отношения «длины» (l) и «высоты» (h) объекта.

Задача 12. Создать выборку размером 100, полученные из так называемого «усеченного» нормального распределения с параметрами $\mathcal{N}(0,1)$ и интервалом определения $[-1,1]$. Результат представить в виде текстового файла.

Задача 13. Какова вероятность попадания выборочных данных за пределы 3σ-интервала (где σ^2 дисперсия распределения) относительно среднего для равномерного распределения?

Задача 14. Имеются наборы измерений длины и ширины прямоугольного участка, содержащие инструментальные ошибки. Если $\{l_i\}$ и $\{w_i\}$ наборы измерени длины и ширины, то оценку площади можно получить, например, двумя способами: $\dfrac{1}{n}\sum_i l_i \cdot w_i$ и $\dfrac{1}{n^2}\sum_i l_i \sum_j w_j$. Провести имитационное моделирование с нормально распределенными величинами с целью установить, какой из подходов наиболее эффективен (оценка площади имеет наименьшую дисперсию).

Задача 15. Выяснить путем имитационного моделирования с выборками из нормального распределения ($\mathcal{N}(0,1)$) величину погрешности при оценивании математического ожидания и дисперсии, привносимую вследствие группировки данных. Предположить, что группировка осуществляется по 10 равным интервалам на инервале $[-10,10]$.