- 3 теоретических занятия. На каждом по 2 темы. По каждой из тем — практическое занятие.
- На следующей паре — консультация.
- После консультации по занятию на сдачу заданий.
- В конце семестра после 6 заданий — самостоятельная (контрольная) работа. Она единственная.
- **НИКАКОГО EXCEL!!!!**
- С си-подобными языками будут проблемы.
- Можно всё посчитать на C, а графики строить в Python.
- Список: Python, C, C++, C#, Java, MatLab, Wolpfram Math, R (доп баллы).
- Программируем алгоритмы не встроенной функции.
- Точки для графика — алгоритмом (вектор)
- Встроенные функции можно использовать для себя.
- Мой вариант — 8.
- Использовать только папку с моим вариантом. Иначе — сразу 0 баллов.
- У других групп — другие данные. За это тоже 0 баллов.

- Каждое задание — 6,5 баллов. Практическая часть (спрашивает что делает строчка кода или подобное) — 3,5 балла. Теоретическая часть — 3 балла. Самостоятельная работа — 10 баллов. Посещаемость — 1 балл. Если не ходить —снимать не будет.
- < 28 баллов — можно добрать баллы. Нужно 50% посещаемости и минимум 3 задания сдать.
- Если не сдали — можно на следующей неделе с половиной баллов можно сдать.
- В каждой папке с вариантом есть Excel файлы. Один — с разделителями-точками, а второй — с запятыми. Это нужно в зависимости от языка программирования.
- В конце папки есть доп.параметры.pdf. там есть нюансы касательно именно различных вариантов.
- Материалы для 1 задания: ст.11-15, ст.51-60.
- Материалы для 2 задания: ст.37-40.

---

1. Описательные статистики
    1. Выборка — n-мерный вектор $x^{(n)} = (x_1, \ldots, x_n)$.
    2. Элементы — случайные величины, такие, что:

    1. $\{x_i\}_{i=1}^{n}$ — независимы в совокупности.
    2. $x_i\ F,\ \forall\, i = \overline{1, n}$.
3. Интерпретации случайной величины:
    - Объект реальной жизни. (погода)
    - Выбор из генеральной совокупности. (озеро с рыбами. Выбираем несколько рыб (вектор рыб). Делаем вывод по этим рыбёшкам и обобщаем на совокупность). Нужен принцип равновероятного выбора.
4. Статистика — любая функция от случайных величин, не зависящая от неизвестных параметров.
5. Базовые статистики:

    1. Выборочное среднее $\overset{\wedge}{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ — оценка $\mathbb{E}X$

    2. Выборочная дисперсия $\overset{\wedge}{\sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overset{\wedge}{\mu})^2$ — оценка $\mathbb{D}X$
    Вспомнили правило трёх сигм + смотрим выборочное стандартное отклонение ($\overset{\wedge}{\delta} = \sqrt{\delta^2}$). Отсюда два случая:

1. $\overset{\wedge}{\sigma}_1 = 25; \quad \overset{\wedge}{\sigma}_2 = 15$ — в среднем с высшим образованием зарабатывают больше.
2. $\overset{\wedge}{\sigma}_1 = 200; \quad \overset{\wedge}{\sigma}_2 = 190$ — основная масса получает приблизительно одну зарплату.

3. Выборочный коэффициент асимметрии.

$$\overset{\wedge}{\gamma}_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overset{\wedge}{\mu})^3}{\overset{\wedge}{\sigma}^3}$$

По нему можно указать, насколько сильно отклонения от среднего значения (средняя зарплата)

4. **Коэффициент эксцесса** (нет в учебнике)

$$\gamma_2 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma^4} - 3$$

$$\overset{\wedge}{\gamma}_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overset{\wedge}{\mu})^4}{\overset{\wedge}{\sigma}^4} - 3$$

5. Выборочная медиана
   $\overset{\wedge}{med} = \hat{Q}(0.5), \hat{Q}(q)$ — выборочный квантиль
   Вариационный ряд выборки $X^{(n)}$ — упорядоченная по неубыванию выборка. Элементы обозначаются с индексом внизу в скобочках $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$

$$\hat{Q}(q) = \begin{cases} X_{((n-1)q+1)}, & \text{если } (n-1)q+1 \text{ — целое} \\ \frac{X_{((n-1)q+1)} + X_{((n-1)q+2)}}{2}, & \text{иначе} \end{cases}$$

6. Гистограмма (частотная и вероятностная). Мы должны построить только вероятностную. Посмотреть подробнее в учебнике. Исследователь сам выбирает, как разделять на интервалы.
   $a_0 = -\infty, \quad a_k = +\infty$
   $a_1 = X_{\min} + \frac{\Delta}{2}, \quad a_{k-1} = X_{\max} - \frac{\Delta}{2}$
   $\Delta = (X_{\max} - X_{\min})k$ (?)
   Формула высоты столбца вероятностной гистограммы: $h_i = \dfrac{\sum\limits_{i=1}^{n} I\big(X_i \in [a_{i-1}, a_i)\big)}{n\Delta}$

   I — индикатор.

Мы по заданию делаем предположение. что данные имеют нормальное распределения. Параметры неизвестны. Мы их оцениваем (выборочное среднее и дисперсия): в функцию плотности нормального распределения подставляем значение параметров (вектор вставляем или ещё чего). Дальше генерируем, например, 100 чисел с одиннаковым интервалом и __. Дальше накладываем на гистограмму.

7. Эмпирическая функция распределения — графическая оценка истинной функции распределения.
   $F(\chi) = \mathbb{P}(X < x)$
   $\hat{F}(\chi) = \dfrac{1}{n}\sum\limits_{i=1}^{n} I(X_i < x)$
   Строим график эмпирической функции распределения. Принимает значения от 0 до 1.
   РИСУНОК.

2. Построение линейной среднеквадратической регрессии (4.2)

   Есть 2 случайные величины x, y. Хотим построить функцию $\hat{y}(x) \approx y$.

   $\hat{y}(x) = ax + b$

   Уравнение среднеквадратической регрессии:

   $$y = \overline{y} + r\frac{\overset{\wedge}{\sigma}_y}{\overset{\wedge}{\sigma}_x}(x - \overline{x}) \qquad (1)$$

   $$x = \overline{x} + r\frac{\overset{\wedge}{\sigma}_x}{\overset{\wedge}{\sigma}_y}(y - \overline{y}) \qquad (2)$$

   r — выборочный коэффициент корреляции.

   Что мы делаем сверхурочно:

   В данных 2 столбца. x, y. Мешать не нужно. Строим диаграмму рассеивания (y от x). Берём первые два значения — ставим точку. И так все координаты распологаем графически. Это и есть диаграмма рассеивания.

   После этого считаем уравнение регрессии (в доп параметрах написано). Для этого считаем коэффичиент корреляции. **Его нужно вывести**. Выводим также средние значения x, y, а также стандартные отклонения x, y.

По данным составляем уравнение регресси. Важно, что мы приводим его к вышеуказанному виду (1, 2).

Далее накладываем график уравнения регрессии на график рассеивания.

Также сделать предсказание (совсем просто). Дан x, скорее всего вне области диаграммы рассеивания. По числу, которое дано, подставляем в формулу и всё.

Часть теоретических вопросов нужно самому отсеить.