Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

Факультет программной инженерии и вычислительной техники

# Отчет

## По лабораторной работе №1

по дисциплине «Математическая статистика»

Вариант 2

Выполнила: Старостина Е. П., группа P3219

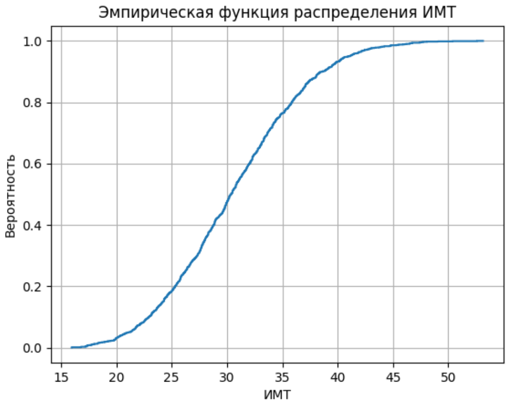
Преподаватель: Лимар И. А.

г. Санкт-Петербург
2025

# Цель работы:

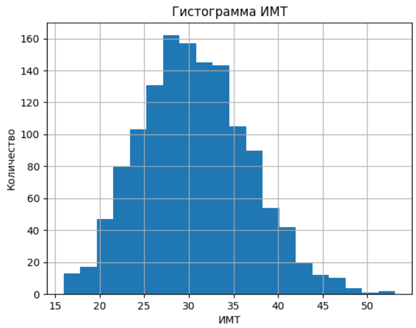
На практике познакомиться с анализом данных, их статистического распределения. Вычислить теоретические характеристики распределения данных.

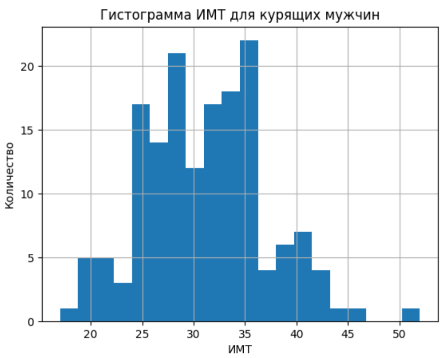
# Результаты задания 1:

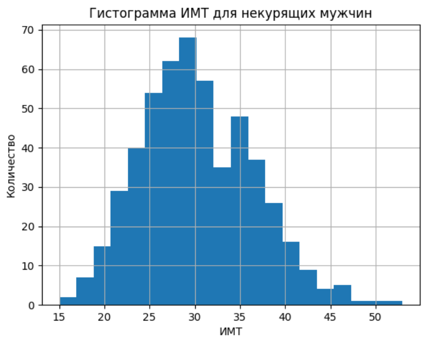


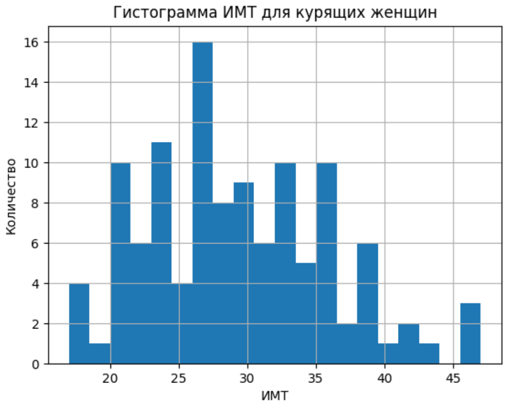


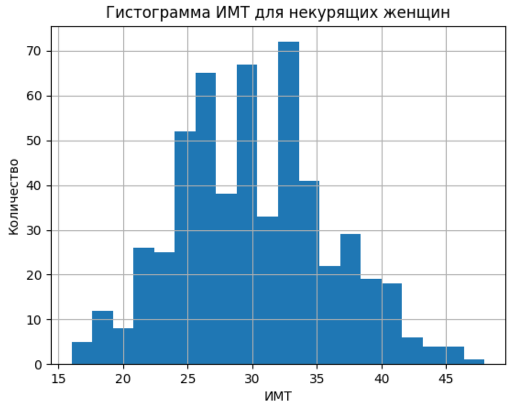


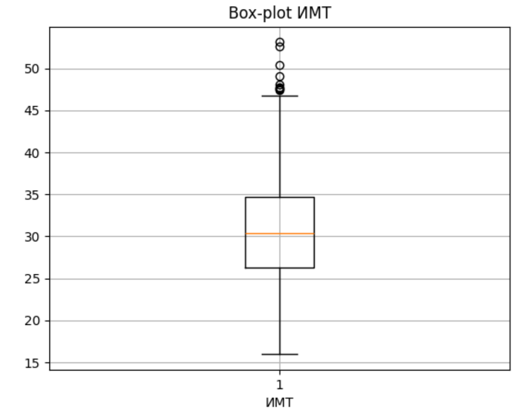



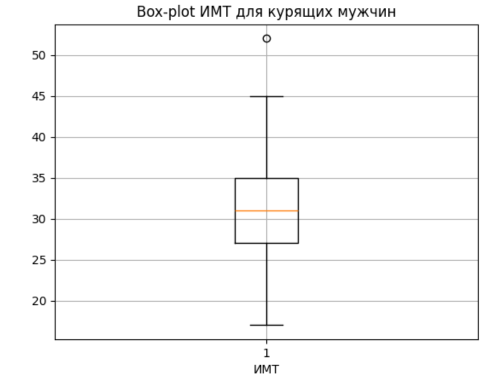



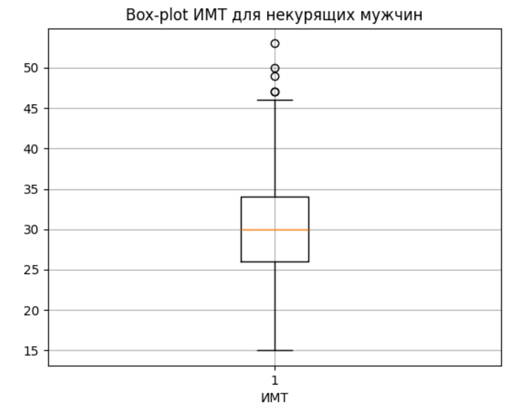



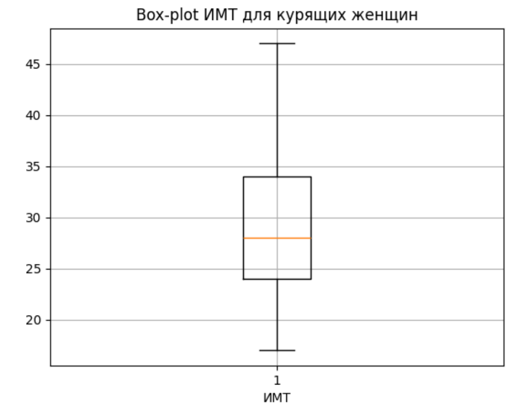

Гистограмма ИМТ для курящих женщин
Количество
ИМТ

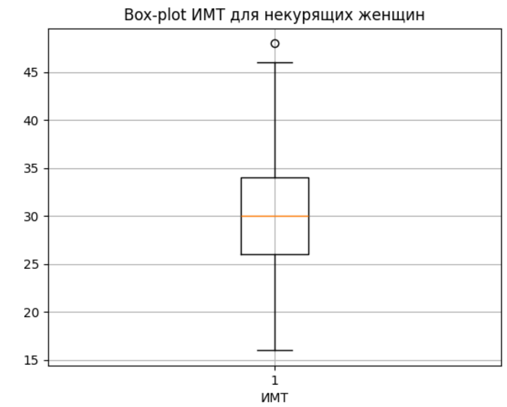

Гистограмма ИМТ для некурящих женщин
Количество
ИМТ


Box-plot ИМТ
ИМТ


Box-plot ИМТ для курящих мужчин
ИМТ


Box-plot ИМТ для некурящих мужчин
ИМТ


Box-plot ИМТ для курящих женщин
ИМТ


Box-plot ИМТ для некурящих женщин
ИМТ

# Задание 2:

Форма гистограммы и эмпирической функции распределения соответствуют нормальному распределению.

Тогда (метод моментов):

$$X \sim N(\mu, \sigma^2)$$

$$E(X) = \mu$$

$$Var(X) = \sigma^2 = E((X - \mu)^2)$$

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\mu = \bar{X}, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Свойства:

1. Согласованность: при $n \to \infty$, выборочное среднее сходится к $\mu$ (закон больших чисел), а $\hat{\sigma}^2$ к $\sigma^2$ ($E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2, \frac{n-1}{n} \to 1$ )
2. Несмещённость: $E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$

Теоретическое смещение:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = (n-1)\sigma^2$$

$$E(\hat{\sigma}^2) = \frac{1}{n}E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2$$

$$Bias(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

Дисперсия:

Если $X_1$, …, $X_n$ – выборка из $N(\mu, \sigma^2)$, то $\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$ имеет распределение с n-1 степенями свободы

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$$

$$Var\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right) = 2(n-1)$$

$$Var(\hat{\sigma}^2) = \left(\frac{\sigma^2}{n}\right)^2 Var\big(\chi^2(n-1)\big) = \left(\frac{\sigma^2}{n}\right)^2 * 2(n-1) = \frac{2\sigma^4(n-1)}{n^2}$$

MSE:

$$MSE(\hat{\sigma}^2) = Var(\hat{\sigma}^2) + Bias(\hat{\sigma}^2)^2 = \frac{2\sigma^4(n-1)}{n^2} + \left(-\frac{\sigma^2}{n}\right)^2 = \frac{2\sigma^4(n-1)+\sigma^4}{n^2}$$
$$= \frac{\sigma^4(2n-1)}{n^2}$$

Информация Фишера:

Функция плотности одного наблюдения:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

Функция правдоподобия:

$$L(\sigma^2) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(X_i-\mu)^2}{2\sigma^2})$$

$$\log\big(L(\sigma^2)\big) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

Первая производная:

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i-\mu)^2$$

Вторая производная:

$$\frac{n}{2\sigma^4} + \frac{1}{2\sigma^6}\sum_{i=1}^{n}(X_i-\mu)^2$$

Мат ожидание:

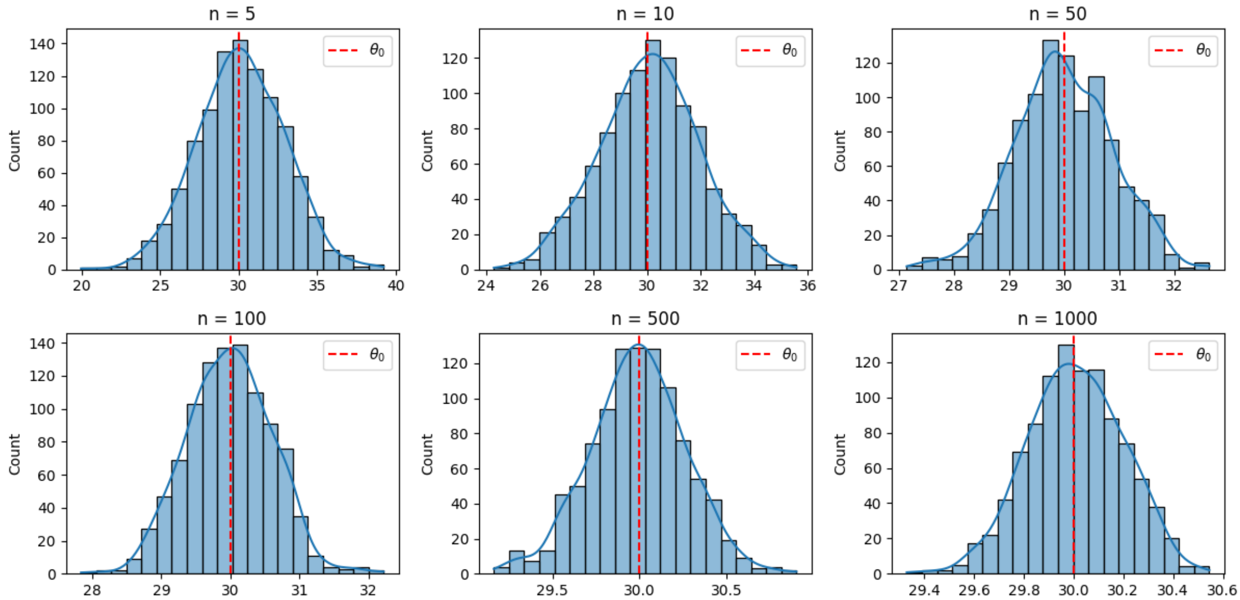
$$E\left(\sum(X_i-\mu)^2\right) = n\sigma^2$$

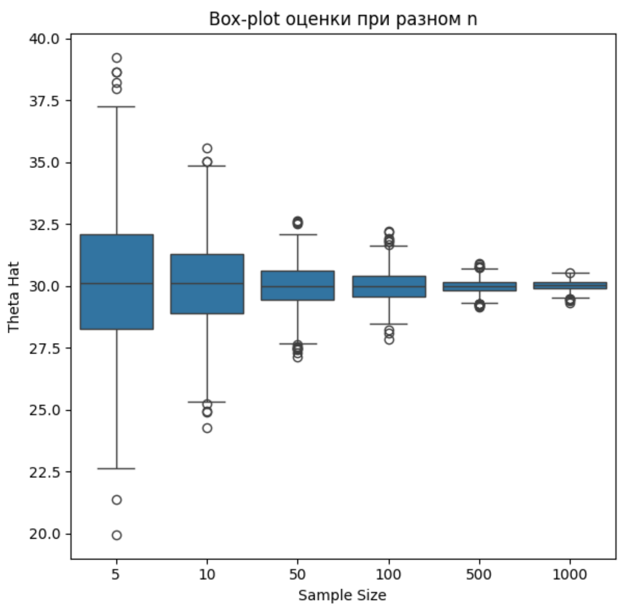
$$E\left(\frac{\partial^2}{\partial(\sigma^2)^2}\log\big(L(\sigma^2)\big)\right) = \frac{n}{2\sigma^4} - \frac{n\sigma^2}{2\sigma^6} = -\frac{n}{2\sigma^4}$$

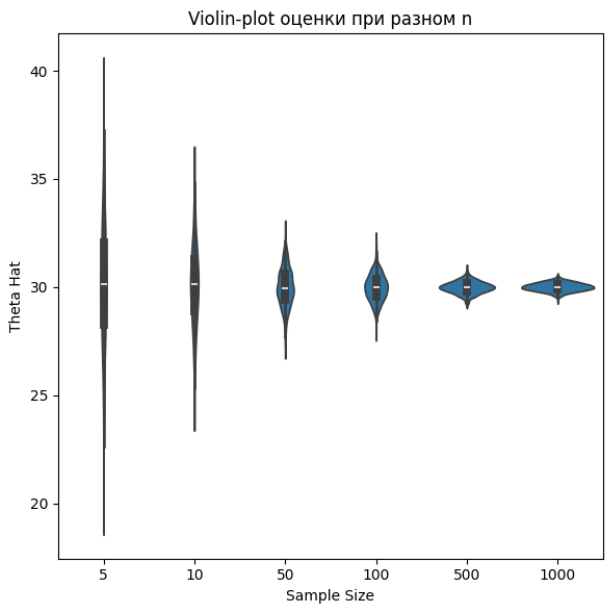
Итого:

$$I(\sigma^2) = \frac{n}{2\sigma^4}$$

## Результаты задания 3:




Как видно из диаграмм, при увеличении n уменьшается разброс оценок, они стремятся к $\theta_0$

```
               count       mean       std  ...       50%       75%
max
Sample Size                                ...

5             1000.0  30.139345  2.782059  ...  30.130734  32.070027
39.219314

10            1000.0  30.086006  1.842079  ...  30.118120  31.303004
35.574506

50            1000.0  30.000409  0.883340  ...  29.969414  30.611450
32.632799

100           1000.0  29.988667  0.613475  ...  29.987080  30.399062
32.215260
```

```
500          1000.0  29.990444  0.274281  ...  29.993003  30.171015
30.902715

1000         1000.0  30.009056  0.191183  ...  30.007904  30.139862
30.545369
```

## Выводы по работе:

В ходе лабораторной работы я на практике познакомилась с анализом данных, их статистического распределения. Вычислила теоретические характеристики нормального распределения данных. Вывела зависимость $\sigma^2$ от объема выборки.

## Приложение:

Ссылка на код:

https://github.com/Starostina-elena/math_stat_lab1