

# Математические основы описания результатов исследования

# Таблица исходных данных

№ п/ п	Испытуемый	Класс	Пол	С/оценка до эксп.	С/оценка после эксп.
1	Аня А.	1	1	80	50
2	Даша Б.	2	1	75	70
3	Витя В.	1	2	60	70
4	Петя Д.	2	2	45	20
5	Ира Ж.	2	1	67	79
6	Коля Н.	1	2	45	73
...	...	...			
$i$	...	$n_i$	$m_i$	$x1_i$	$x2_i$
...	...	...			
60	Женя Я.	1	2	35	75

# Группировка данных

**Группировка** — процесс систематизации и упорядочивания данных с целью извлечения содержащейся в них информации

Способы группировки:

- А) **безинтервальный вариационный ряд;**
- Б) **интервальный вариационный ряд.**

Данные могут быть представлены в аналитической и графической форме

# Вариационный ряд: термины

**Вариационный ряд** — ранжированный в порядке возрастания/убывания ряд значений признака с соответствующими им весами (количествами людей, обладающих этим уровнем признака).

1) **Дискретный вариационный ряд** — частоты привязаны к конкретным значениям признака.

Пример: у 5 детей по тесту школьной готовности 10 баллов; 3 ребенка получили 7 баллов; 4 ребенка — 8 баллов; и один ребенок — 5 баллов.

# Таблица распределения частот

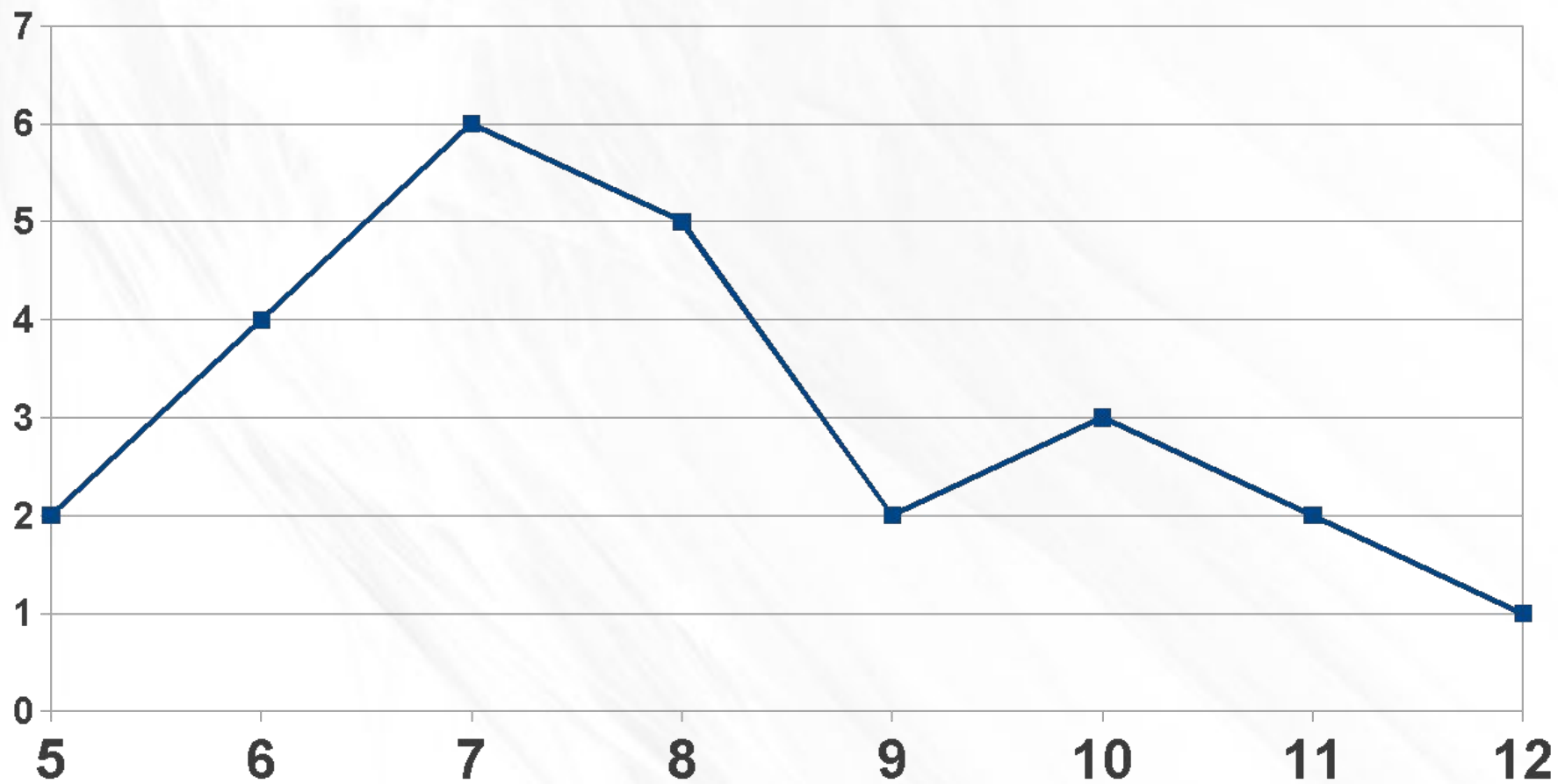
Тест Керна-Иерасека, набор в 1 «Д» класс ( $N=25$  чел)

$x_i$	5	6	7	8	9	10	11	12	Балл по тесту
	++	++++	++++ ++	++++ +	++	+++	?	+	Разнесение выборки
$m_i$	2	4	6	5	?	?	2	1	Частота
$P_i$	0,08	0,16	0,24	0,20	0,08	0,12	0,08	?	Относит. частота
$M_{xi}$	2	6	12	17	19	22	24	25	Накопл. частота
$P_{xi}$	0,08	0,24	0,48	0,68	0,76	0,88	0,96	1,00	Накопл. относит. частота

# Определения

- **Частота** - число, показывающее, сколько раз данное значение признака встречается в данной выборке.
- **Относительная частота** — отношение частоты к объему выборки
- **Накопленная частота** — сумма частот от первого имеющегося значения до текущего
- **Накопленная относительная частота** — аналогично см. выше

# Полигон частот



# Интервальный вариационный ряд

В **интервальном вариационном ряду** частоты, характеризующие повторяемость вариантов в выборке, распределяются по интервалам группировки;

Применяется, если:

- а) есть большой размах вариации;
- б) объем выборки большой;
- в) измеряли непрерывную величину (время, скорость и т.п.)



# Интервальный вариационный ряд

- **Количество интервалов** - по формуле Стерджеса, в которой **k** - число интервалов, **N**-объем выборки

$$k=1+3,32*\lg N$$

Количество интервалов - по таблице:

<b>N</b>	<b>k</b>
25-40	5-6
40-60	6-8
60-100	7-10
100-200	8-12
>200	10-15

# Интервальный вариационный ряд

- **Ширина интервала** определяется по формуле:

$$h = (X_{\max} - X_{\min}) / k$$

Есть два правила:

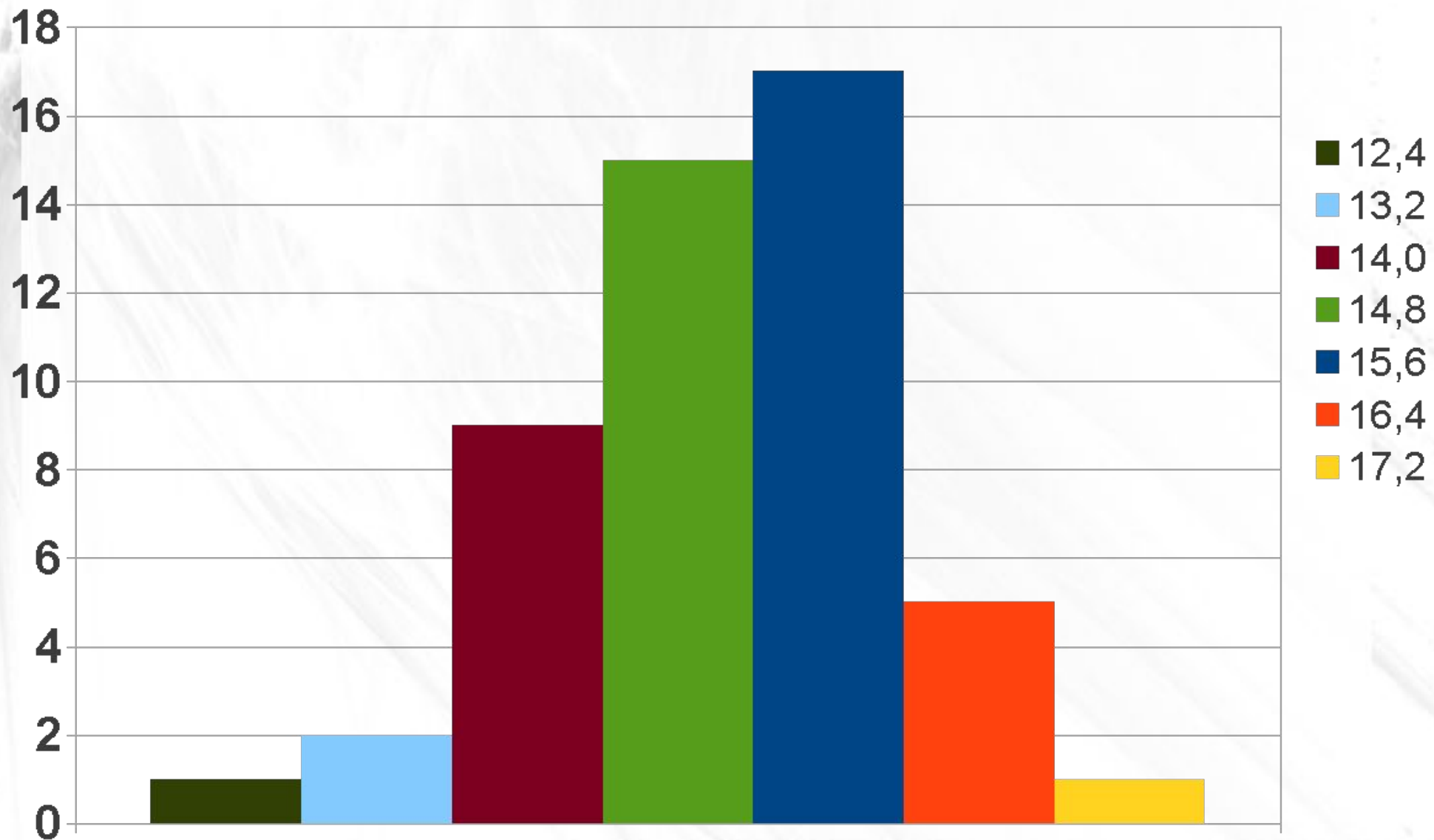
**Правило 1.** Точность границ интервала должна быть равна точности данных (число знаков после запятой одинаково)

**Правило 2.** Минимальное значение должно попадать в середину 1-го интервала

# Интервальный вариационный ряд

№ п/п	Границы	Разнесение	Середина	Частота	Отн. частота
1	12,4-13,2	+	12,8	1	0,02
2	13,2-14,0	++	13,6	2	0,04
3	14,0-14,8		14,4	9	0,18
4	14,8-15,6		15,2		
5	15,6-16,4		16,0		
6	16,4-17,2		16,8		
7	17,2-18,0		17,6		
Итого:		50	-	50	1,00

# Гистограмма распределения частот



# Гистограмма

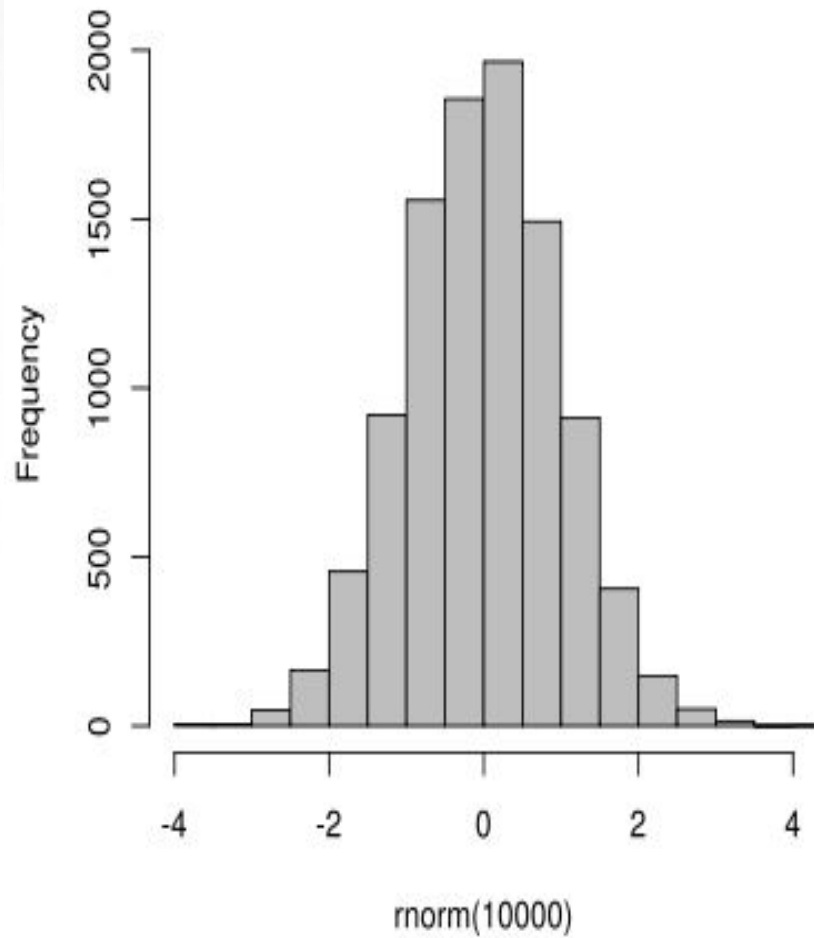
**Гистограмма** — график, отображающий особенности интервального ряда распределения (термин ввел К.Пирсон в 1895г.)

- Основанием столбика диаграммы является интервал группировки
- При построении гистограммы, описывающей плотность распределения, высота  $i$ -го столбика ( $L$ ) зависит от частоты ( $m$ ) данного интервала:

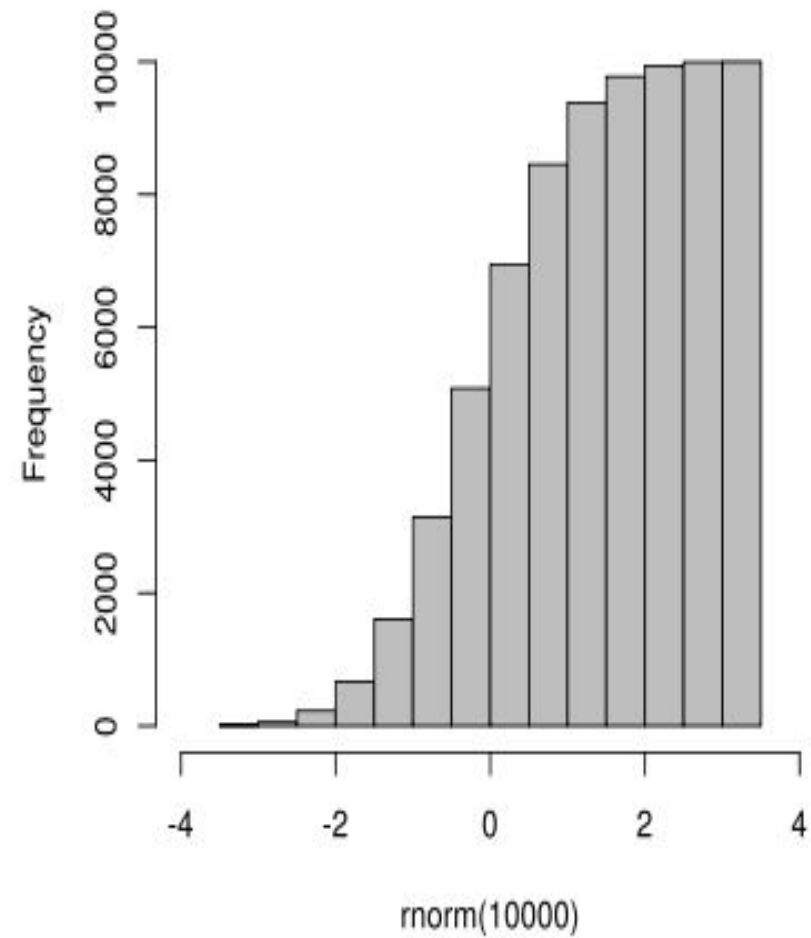
$$L_i = m_i / (h * N)$$

# Виды гистограмм

Ordinary histogram



Cumulative histogram



# Для чего нужны гистограммы?

- ✓ Понять истинную природу различий двух выборок

Пример:

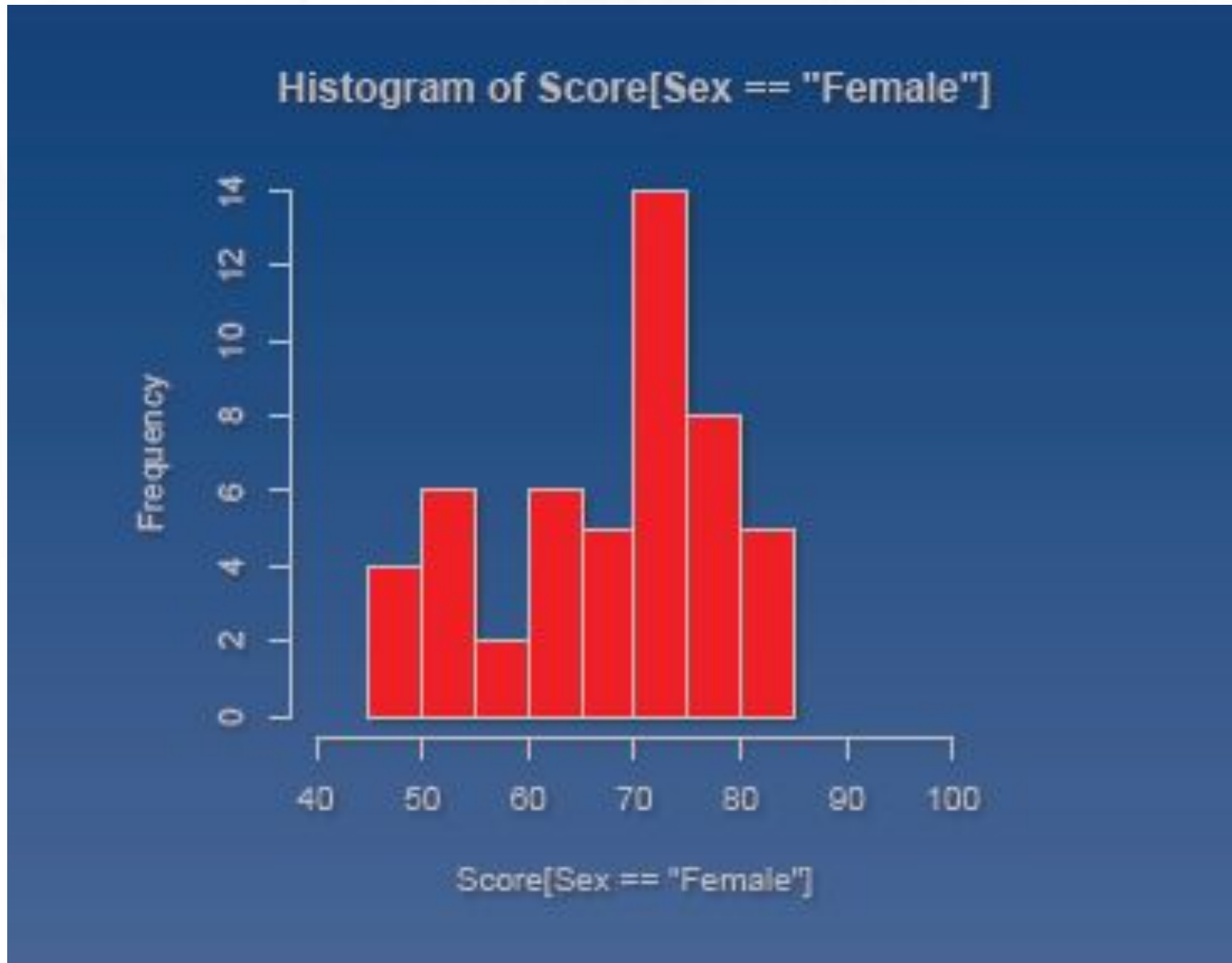
50 мужчин и 50 женщин;

1 тест пространственного мышления (0-100);

**Мужчины:** ср.арифм=**72,62**

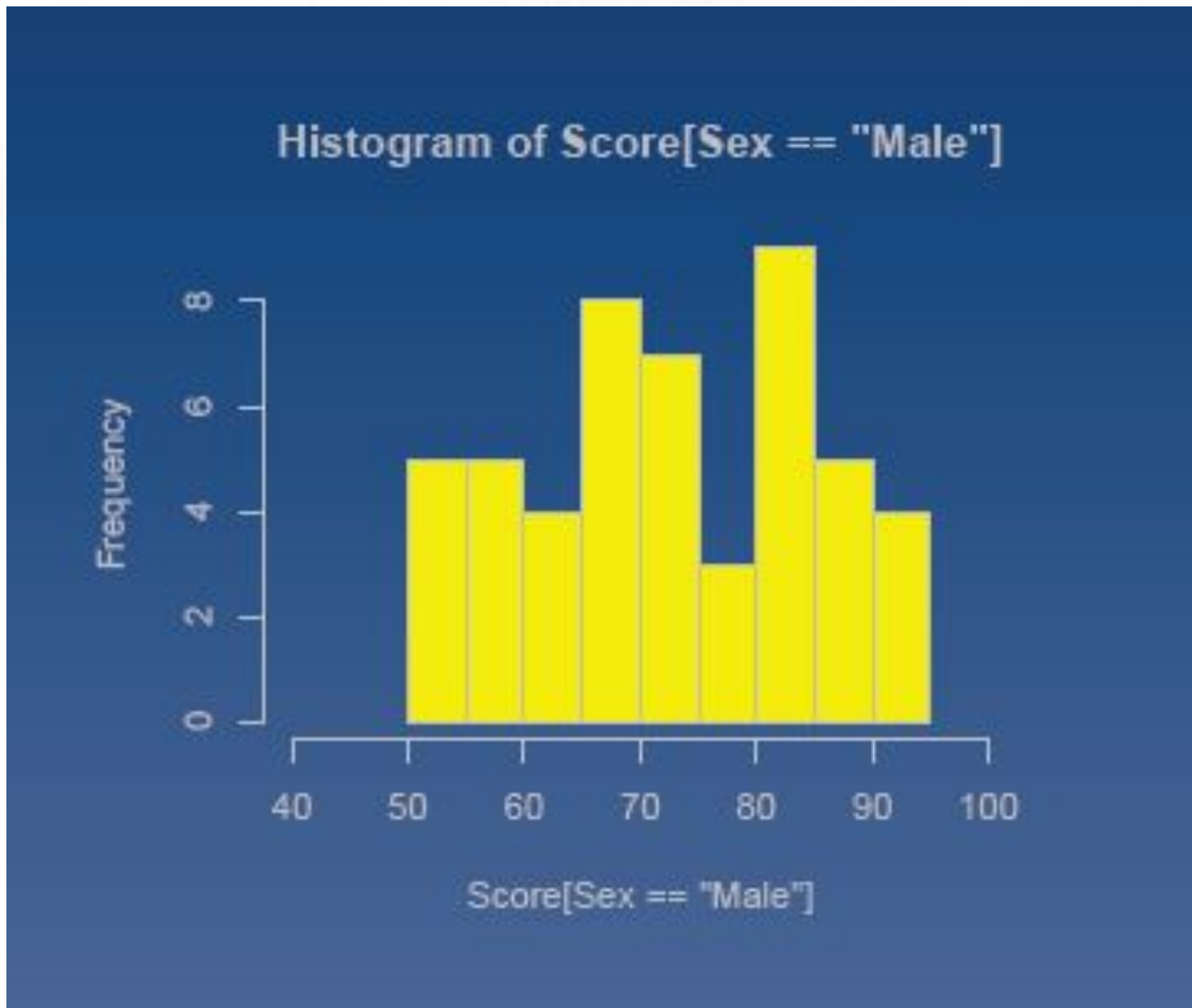
**Женщины:** ср.арифм=**67,92**

# Пример: женщины

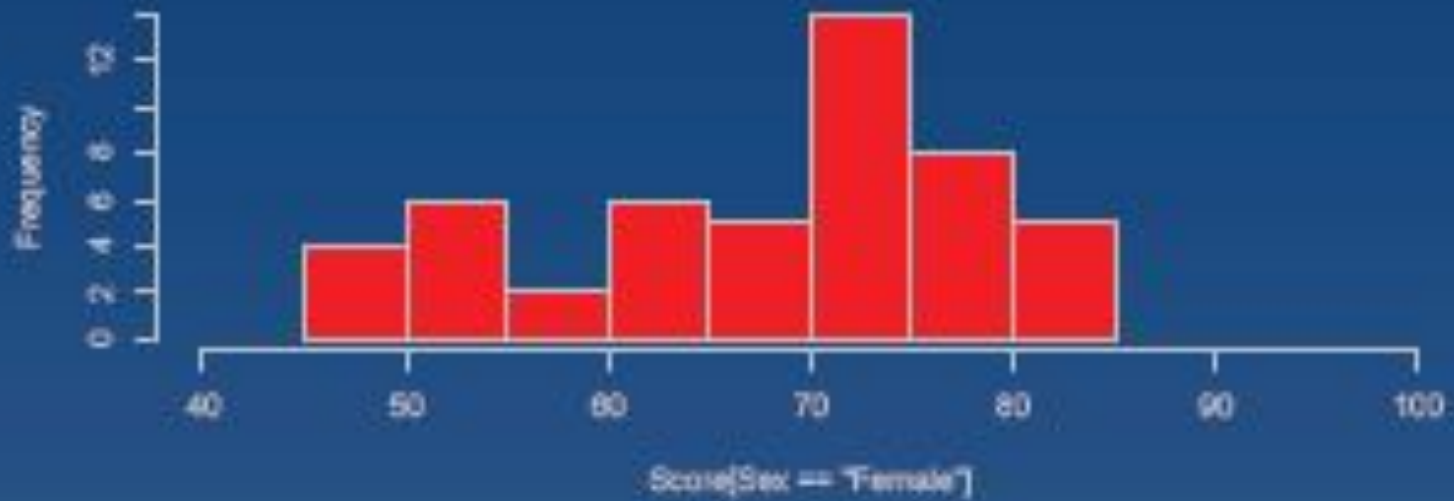




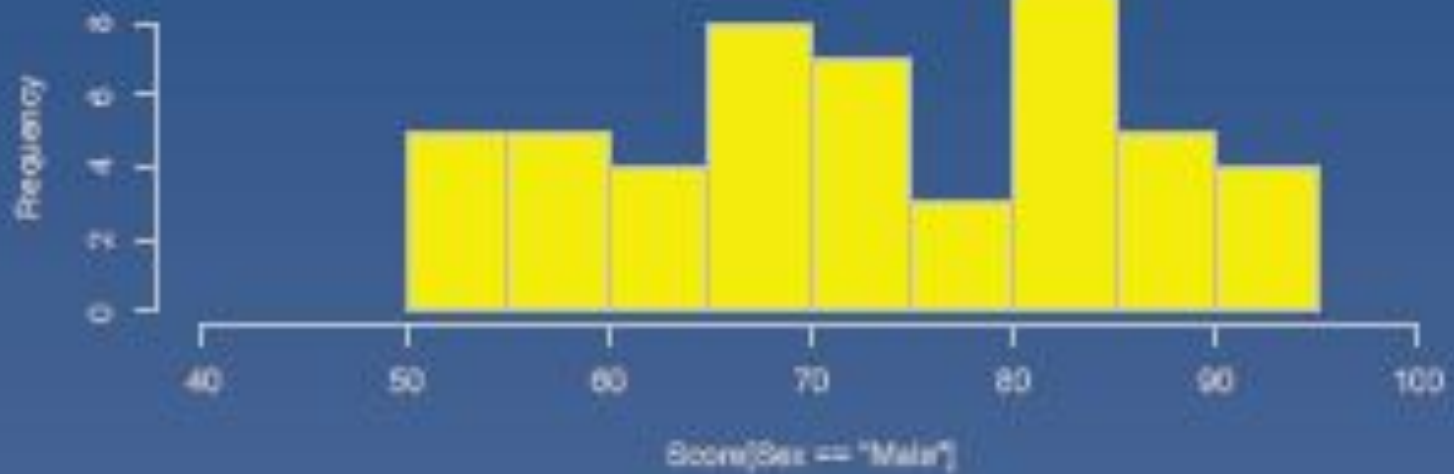
# Пример: мужчины



Histogram of Score[Sex == "Female"]



Histogram of Score[Sex == "Male"]



# Для чего нужны гистограммы?

- ✓ Понять, есть ли **скрытые переменные**

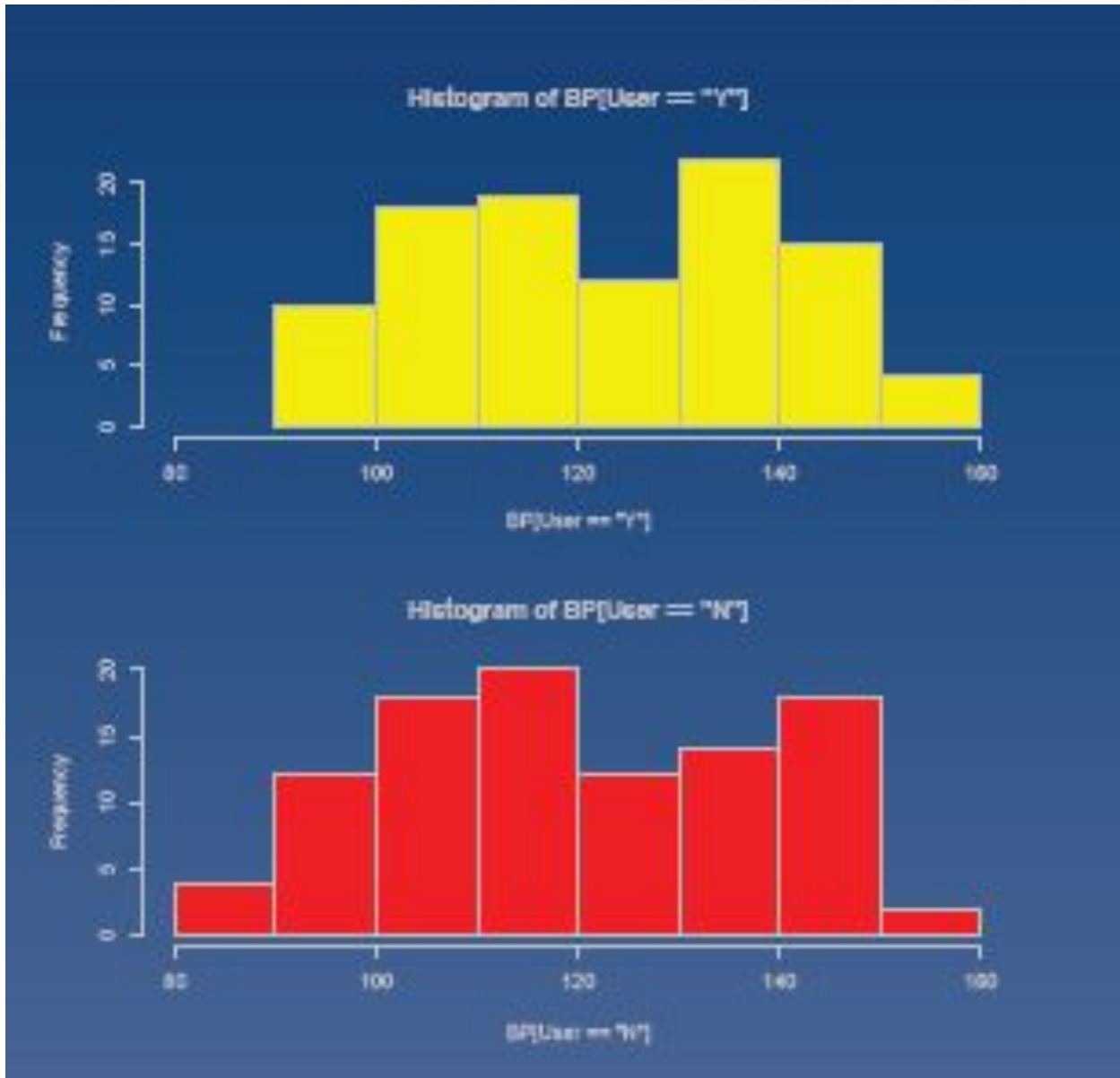
Пример:

Влияет ли оральная контрацепция на уровень Артериального Давления (АД)?

100 женщин, использующих “The Pill”

100 женин, использующих другие виды контрацепции

# Сравним гистограммы:



Используют  
оральную  
контрацепцию

АД=124,41

**Используют  
другие виды**

**АД=119,85**

# Задание

**Возраст рабочих** цеха (в годах):

18; 38; 40; 28; 29; 26; 38; 34; 22; 28; 30; 22; 23;  
35; 33; 27; 24; 30; 32; 49; 37; 28; 25; 29; 26;  
31; 24; 29; 27; 32; 25; 29; 29; 52; 58; 44; 39;  
57; 19; 25.

Для анализа распределения рабочих цеха по возрасту требуется:

1. Построить интервальный ряд распределения;
2. Дать графическое изображение ряда.  
Сформулировать вывод.