

## Методы статистической обработки информации - Задание 1

Ершов А. С., гр. 22.М04-мм

Задание: по реальным данным найти среднее, медиану, моду, дисперсию, рассеяние, асимметрию, эксцесс.

Table: Распределения числа аномалий на ядре.

Доза, Гр	in vitro						in vivo					
	число аномалий						число аномалий					
	0	1	2	3	4	> 5	0	1	2	3	4	> 5
0	68	25	7				66	31	2	1		
5	74	19	5	2			50	35	13	2		
10	59	24	16	1			41	39	17	2	1	
15	48	33	11	3	2	3	27	39	29	3	1	1
20	59	31	5	4	1		22	22	32	15	6	3
25	37	31	22	6	2	2	33	39	18	8	1	1
30	35	37	17	5	3	3	21	29	21	14	10	5
35	26	36	19	10	5	4	17	24	32	11	10	6
40	19	33	25	11	7	5	15	17	29	18	11	10
45							13	14	24	24	12	13

Мой вариант: in vitro, 35 Гр.

Выполнено на языке R в R Notebook.

Создадим выборку по имеющимся данным.

```
data <- rep(c(0,1,2,3,4,5),times=c(26,36,19,10,5,4))
data
```

```
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
## [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
## [75] 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 5 5 5 5
```

Рассчитаем заданные величины.

Среднее:

```
result.mean <- round(mean(data), 1)
print(result.mean)
```

```
## [1] 1.4
```

*Медиана:*

```
result.median <- round(median(data), 1)
print(result.median)
```

```
## [1] 1
```

*Мода:*

```
result.mode <- unique(data)
result.mode <- result.mode[which.max(tabulate(match(data, result.mode)))]
print(result.mode)
```

```
## [1] 1
```

Наиболее часто встречающийся в выборке элемент - 1, по условию он встречается 36 раз.

*Дисперсия:*

```
result.var <- round(var(data), 1)
print(result.var)
```

```
## [1] 1.8
```

*Асимметрия:*

```
library(moments)
result.skewness <- round(skewness(data), 1)
print(result.skewness)
```

```
## [1] 1
```

Асимметрия получилась больше 0, то есть “правый хвост” распределения длиннее левого, правосторонняя асимметрия.

*Экссесс:*

```
result.kurtosis <- round(kurtosis(data), 1)
print(result.kurtosis)
```

```
## [1] 3.4
```

Экссесс получился больше 0, хвосты распределения “легче”, а пик острее, чем у нормального распределения.

*Рассеяние:*

```
result.scattering <- round(result.var / result.mean, 1)
print(result.scattering)
```

```
## [1] 1.3
```

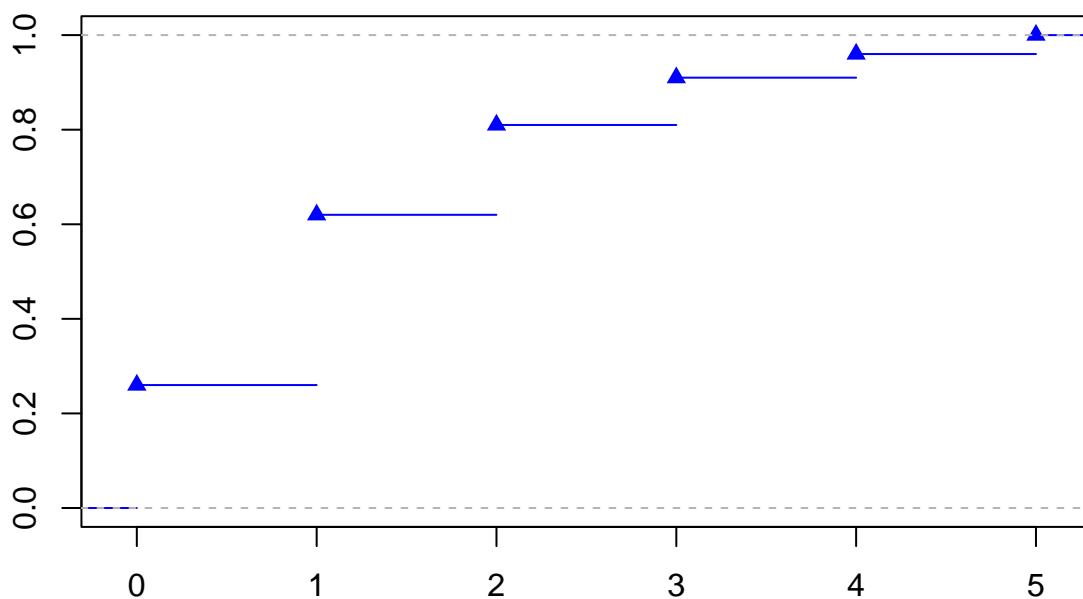
Сведём результаты в таблицу:

```
result_df <- data.frame(result.mean, result.median, result.mode, result.var, result.skewness, result.kurtosis)
colnames(result_df) <- c('Среднее', 'Медиана', 'Мода', 'Дисперсия', 'Асимметрия', 'Экссесс', 'Рассеяние')
print(result_df)
```

```
## Среднее Медиана Мода Дисперсия Асимметрия Экссесс Рассеяние
## 1      1.4      1      1      1.8      1      3.4      1.3
```

Построим график функции эмпирического распределения:

```
ecdf_res <- ecdf(data)
plot(ecdf_res, xlim=c(-0.1, 5.1), pch = 17, main="", xlab="", ylab="", col="blue")
```



Построим график относительных частот.

```
relative_freq <- c(26,36,19,10,5,4) / 100
barplot(relative_freq,
        names.arg = c("0", "1", "2", "3", "4", "5"),
        xlab = "Anomalies",
        ylab = "Relative frequency")
```

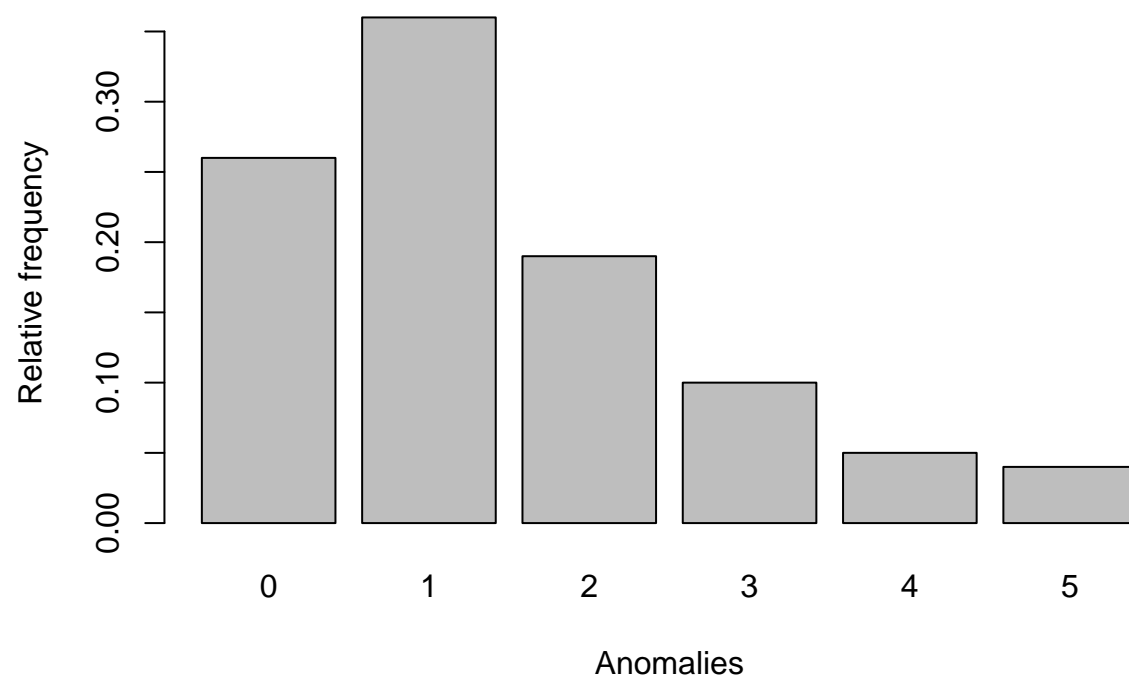


График подтверждает вывод относительно эксцесса и асимметрии.