Теория вероятностей, ДЗ №3

Мироненко Ольга

Задание 1

Количество циклов - это дискретная случайная величина.

Обозначим количество циклов за x, а вероятность соответствующего количества циклов за p, тогда математическое ожидание для количества циклов может быть рассчитано по формуле:

$$M(x) = \sum_{i=1}^{N} x_i * p_i$$

где N - количество уникальных значений дискретной случайной величины (в нашем случае $N = 2$ и для первичных случаев, и для рецидивов). Дисперсия может быть оценена по формуле:

$$D(x) = \sum_{i=1}^{N} (x_i - M(x))^2 * p_i$$

Применив эти формулы для таблиц распределений количества циклов, получим:

- по первичным случаям:
  - математическое ожидание: $M(x) = 1 * 0.5 + 2 * 0.5$ = 1.5 цикла
  - дисперсия: $D(x) = (1 - 1.5)^2 * 0.5 + (2 - 1.5)^2 * 0.5$ = 0.25 цикла^2
- по рецидивам:
  - математическое ожидание: $M(x) = 2 * 0.25 + 3 * 0.75$ = 2.75 цикла
  - дисперсия: $D(x) = (2 - 2.75)^2 * 0.25 + (3 - 2.75)^2 * 0.75$ = 0.1875 цикла^2

Задание 2

В случае рецидива заболевания общее количество циклов терапии, которые может получить пациент, складывается из циклов, полученных при первичной терапии, и циклов, полученных в связи с рецидивом. Таким образом, учитывая условия задачи, можем сказать, что общее количество циклов терапии для пациента с рецидивом - это случайная величина, равная сумме двух независимых случайных величин. Можем изобразить таблицу распределения в следующем виде:

Общее кол-во циклов	1+2	1+3	2+2	2+3
Вероятность	$p_{01} * p_{12}$	$p_{02} * p_{13}$	$p_{02} * p_{12}$	$p_{02} * p_{13}$

где p_{ij} - это вероятность случайной величины i ($i = 0$ для первичных случаев, $i = 1$ - для рецидивов) принять значение j.

Общее количество циклов терапии для пациентов с рецидивом, таким образом, может принимать 3 уникальных значения: 3, 4 и 5, поэтому итоговая таблица распределения будет выглядеть следующим образом:

Общее кол-во циклов	3	4	5
Вероятность	$p_{01} * p_{12}$	$p_{02} * p_{13} + p_{02} * p_{12}$	$p_{02} * p_{13}$

Общее кол-во циклов	3	4	5
Вероятность	$0.5 * 0.25$	$0.5 * 0.75 + 0.5 * 0.25$	$0.5 * 0.75$

Общее кол-во циклов	3	4	5
Вероятность	0.125	0.5	0.375

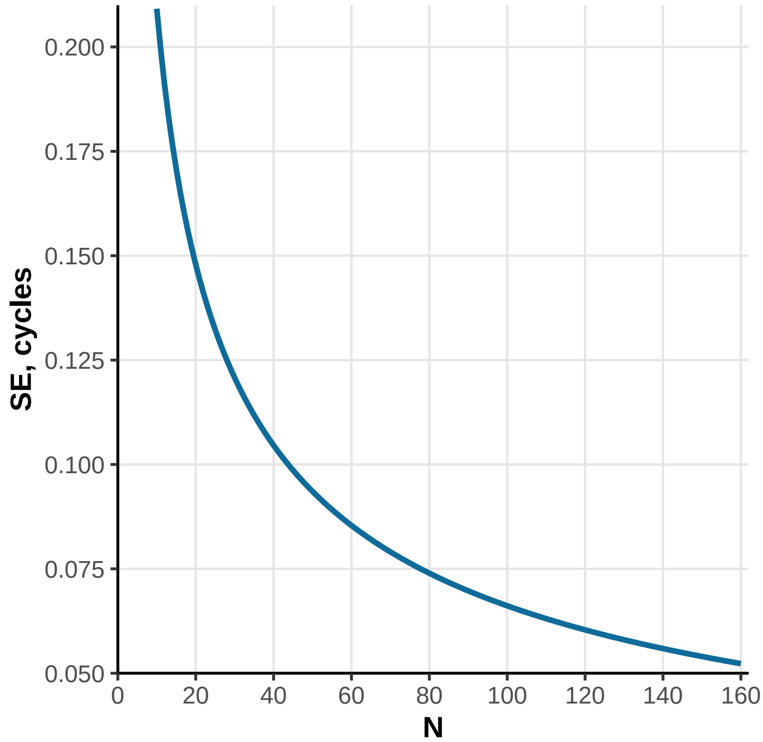
Математическое ожидание этой случайной величины: $M(x) = 3 * 0.125 + 4 * 0.5 + 5 * 0.375 = 4.25$ цикла.

Дисперсия: $D(x) = (3 - 4.25)^2 * 0.125 + (4 - 4.25)^2 * 0.5 + (5 - 4.25)^2 * 0.375 = 0.4375$ цикла^2.

Задание 3

Стандартная ошибка среднего может быть рассчитана по формуле: $SE = \frac{\sqrt{D(x)}}{\sqrt{N}}$.

График зависимости стандартной ошибки среднего числа циклов терапии для пациента с рецидивом от размера выборки:



Задание 4

```
# истинное значение дисперсии (в ген.совокупности)
var_th <- (3-4.25)^2*0.125+(4-4.25)^2*0.5+(5-4.25)^2*0.375

set.seed(123)

df <- tibble(n = c(10, 40, 160)) %>%
  rowwise() %>%
  mutate(
    # случайная выборка из заданного распределения числа циклов
    x = list(sample(x = c(3,4,5), size = n,
                    replace = TRUE, prob = c(0.125,0.5,0.375))),
    # выборочная оценка стандартного отклонения
    sd_hat = sd(x),
    # выборочная оценка стандартной ошибки среднего
    se_hat = sd_hat/sqrt(n),
    # теоретическая стандартная ошибка среднего
    se = sqrt(var_th/n),
    # разница между выборочной оценкой и
    # теоретическим значением стандартной ошибки
    se_diff = se_hat - se) %>%
  mutate_at(vars(sd_hat, contains("se")), ~round(., 4))

knitr::kable(df %>% select(-x),
             col.names = c("n", "$\\hat{\\sigma}$", "$\\hat{SE}$",
                           "$SE$", "$\\hat{SE}-SE$"),
             align = "c", escape = FALSE) %>%
  kableExtra::kable_styling(latex_options = "HOLD_position")
```

n	$\hat{\sigma}$	$\hat{SE}$	SE	$\hat{SE}-SE$
10	0.8165	0.2582	0.2092	0.0490
40	0.6864	0.1085	0.1046	0.0039
160	0.6490	0.0513	0.0523	-0.0010

С увеличением объёма выборки оценка стандартной ошибки среднего числа циклов меньше отклоняется от теоретического значения. Это связано с тем, что с увеличением объёма выборки выборочная оценка дисперсии приближается к дисперсии в генеральной совокупности.