# Определение минимального количества объектов, необходимого для проведения выборочного эксперимента*

Л. Г. Садыхова, Д. Б. Назаренко, И. А. Бабаев

**Аннотация.** На основе сравнения характеристик надежности отказавших и неотказавших объектов в течение заданного времени определяется минимальное количество объектов, необходимое для проведения выборочного эксперимента.

**Ключевые слова:** *выборка, наработка до отказа, средний ресурс, доверительная вероятность, точечная оценка, нижняя доверительная граница, средняя доля безотказных наработок.*

## Введение

При планировании выборочного эксперимента на заданное время следует определить минимальный размер выборки — количество однотипных объектов, необходимое для проведения такого рода исследований.

Актуальность поставленной авторами задачи обусловлена, в частности, универсальной востребованностью данного типа эксперимента в любой сфере знания — от технических наук до социальных [1-3].

В целях определения искомой величины воспользуемся терминологией, сложившейся в планировании испытаний технических объектов на надежность.

Пусть $\tau$ — наработка до отказа невосстанавливаемого объекта.

Введем следующую смешанную случайную величину на отрезке времени $(0,t)$:

$$\eta(t)=\begin{cases}\tau, \text{ если } \tau<t;\\ t, \text{ иначе.}\end{cases}$$

Следовательно, величина $\eta(t)$ — безотказная наработка объекта в течение времени $t$.

Определим среднюю долю безотказной наработки (далее СДБН) объекта в течение времени $t$ по следующей формуле:

$$\Upsilon(t)=M\left(\frac{\eta(t)}{t}\right), \tag{1}$$

где $M(\cdot)$ — математическое ожидание величины, стоящей внутри скобок.

Другими словами, СДБН — это математическое ожидание отношения безотказной наработки объекта в течение периода наблюдения за объектом к продолжительности этого периода.

Заметим, что величина

$$\rho(t)=M\left(\eta(t)\right)$$

— средняя безотказная наработка объекта в течение времени $t$. Легко заметить, что

$$\Upsilon(t)=\frac{\rho(t)}{t},$$

т. е. показатель СДБН — это доля средней безотказной наработки (ДСБН) объекта в течение периода наблюдения за ним (т. е. СДБН = ДСБН).

Очевидно, что

$$\bar{\Upsilon}(t)=1-\Upsilon(t) \tag{2}$$

— средняя доля невыработанной наработки объекта в течение промежутка времени $t$.

Чтобы дать еще одну интерпретацию показателя СДБН, докажем следующее утверждение:

*Теорема 1*. Справедлива следующая формула:

$$\Upsilon(t)=\frac{1}{t}\int_0^t P(x)dx, \tag{3}$$

где $P(x)$ — вероятность безотказной работы объекта в течение времени $x$.

* Работа выполнена при финансовой поддержке РФФИ (проект № 10–08–00607–а).

*Доказательство*. Пусть $f_\eta(x)$ — плотность распределения случайной величины $\eta(t)$. Тогда

$$f_\eta(x)=\begin{cases} f(x), \text{ если } x\in(0,t); \\ P(t), \text{ если } \quad x=t, \end{cases}$$

где $f(x)$ — плотность распределения случайной величины $\tau$.

Согласно формуле математического ожидания смешанной случайной величины $\eta(t)$ имеем

$$M\left(\eta(t)\right)=tP(t)+\int_0^t xf(x)dx\,, \tag{4}$$

откуда с учетом того, что

$$\int_0^t xf(x)dx=-tP(t)+\int_0^t P(x)dx\,,$$

и принимая во внимание (1), получим искомую формулу (3).

Из формулы (3) следует, что показатель СДБН — это среднее значение вероятности безотказной работы объекта в течение времени $t$.

Формула (3) позволяет доказать следующие свойства показателя СДБН:

1. $0\le \Upsilon(t)\le 1$;
2. $\lim\limits_{t\to+0}\Upsilon(t)=1$;
3. $\lim\limits_{t\to\infty}\Upsilon(t)=0$ при условии, что ограничена средняя наработка до отказа;
4. $\Upsilon(t)\ge P(t)$;
5. $\Upsilon'(t)=\frac{1}{t}\left(P(t)-\Upsilon(t)\right)$;
6. $\Upsilon(t)-P(t)=\frac{1}{t}\int_0^t xf(x)dx$.

Первые два свойства очевидны.

Третье свойство следует из того, что

$$\lim_{t\to\infty}\int_0^t P(x)dx=r<\infty\,,$$

где $r$ — средняя наработка до отказа.

Четвертое свойство вытекает из следующей оценки:

$$\int_0^t P(x)dx\ge tP(t)\,.$$

Для доказательства пятого свойства воспользуемся тем, что

$$t\Upsilon(t)=\int_0^t P(x)dx\,,$$

откуда, взяв производную по времени $t$, легко получим искомое свойство.

Шестое свойство вытекает из соотношения (4) и формулы (1).

## 1. Точечная оценка показателя СДБН

Пусть испытываются на безотказность $n$ однотипных объектов, из которых $k$ объектов отказало в течение времени испытания $t$. Обозначим через $\tau_i$ безотказную наработку отказавшего $i$-го объекта ($i=1,2,...,k$). Тогда точечной оценкой показателя $\Upsilon(t)$ служит величина

$$\widehat{\Upsilon}_n(t)=\frac{1}{nt}\left(\sum_{i=1}^k\tau_i+(n-k)t\right). \tag{5}$$

В самом деле, т. к. $\tau_i/t$ — доля наработки до отказа $i$-го отказавшего объекта в течение периода наблюдения $t$, а полная доля безотказной наработки неотказавшего объекта в течение того же периода наблюдения равна единице, то согласно (1) точечная оценка показателя $\Upsilon(t)$ равна

$$\widehat{\Upsilon}_n(t)=\frac{\sum_{i=1}^k\tau_i/t+(n-k)}{n}\,, \tag{6}$$

что и доказывает формулу (5).

Из формулы (5) следует, что точечная оценка показателя (2) рассчитывается по формуле

$$\widehat{\widehat{\Upsilon}}_n(t)=\frac{1}{nt}\sum_{i=1}^k\left(t-\tau_i\right). \tag{7}$$

Формулы (5) и (6) наглядным образом отражают свойства 1, 4, и 6 для показателя $\Upsilon(t)$, где

$$\widehat{P}_n(t)=\frac{n-k}{n} \tag{8}$$

— точечная оценка вероятности безотказной работы объекта в течение времени $t$.

В оценке (5) величины $k$ и $\tau_i$ ($i=1,2,...,k$) случайны. Поэтому возникает вопрос: смещена ли полученная оценка? В связи с этим докажем следующее утверждение:

*Теорема 2*. Точечная оценка $\widehat{\Upsilon}_n(t)$, определенная формулой (5) для показателя $\Upsilon(t)$, несмещенная, т. е. справедлива следующая формула:

$$M\left(\widehat{\Upsilon}_n(t)\right)=\Upsilon(t)\,. \tag{9}$$

*Доказательство*. Из формулы (6) находим

$$\hat{\Upsilon}_n(t)-\widehat{P_n}(t)=\frac{\sum_{i=1}^{k}\tau_i/t}{n},$$

где $\widehat{P_n}(t)$ — точечная оценка вероятности безотказной работы объекта в течение времени $t$, определенная формулой (8).

Откуда имеем

$$M\left(\hat{\Upsilon}_n(t)-\widehat{P_n}(t)\right)=M\left(\frac{\sum_{i=1}^{k}\tau_i}{nt}\right). \qquad (10)$$

Для правой части получим

$$M\left(\frac{\sum_{i=1}^{k}\tau_i}{nt}\right)=\frac{1}{nt}M\left(\sum_{i=1}^{k}\tau_i\right). \qquad (11)$$

Будем считать, что наработки отказавших объектов на отрезке времени $(0,t)$ имеют одну и ту же функцию распределения $F_t(x)$. Найдем ее.

Для условной вероятности при $x<t$ имеем, согласно теореме умножения вероятностей, следующее соотношение:

$$\Pr\left((\tau<x)/(\tau<t)\right)=\frac{\Pr\left((\tau<x)\cap(\tau<t)\right)}{\Pr(\tau<t)},$$

где $\Pr(\cdot)$ — вероятность события, заключенного в скобках. Откуда находим

$$F_t(x)=\frac{F(x)}{F(t)},$$

где $F(\cdot)$ — функция распределения наработок до отказа исследуемого объекта.

Следовательно, для математического ожидания наработки до отказа внутри отрезка времени $(0,t)$ имеем

$$M(\tau)=\frac{1}{F(t)}\int_0^t xf(x)dx, \qquad (12)$$

где $f(x)=F'(x)$ — плотность распределения наработок до отказа на всей временной оси.

Далее, согласно тождеству Вальдо [4], имеем

$$M\left(\sum_{i=1}^{k}\tau_i\right)=M(k)\cdot M(\tau),$$

где, по формуле Бернулли,

$$M(k)=nF(t).$$

Тогда, учитывая (12) после упрощения, получим

$$\frac{1}{nt}M\left(\sum_{i=1}^{k}\tau_i\right)=\frac{1}{t}\int_0^t xf(x)dx.$$

Следовательно, согласно (10) с учетом (11), имеем

$$M\left(\hat{\Upsilon}_n(t)-\widehat{P_n}(t)\right)=\frac{1}{t}\int_0^t xf(x)dx.$$

Откуда, с учетом того, что

$$M\left(\widehat{P_n}(t)\right)=P(t),$$

получим

$$M\left(\hat{\Upsilon}_n(t)\right)-P(t)=\frac{1}{t}\int_0^t xf(x)dx.$$

Сравнивая эту формулу с формулой свойства (6) для показателя $\Upsilon(t)$, получаем искомое соотношение (9), что и доказывает теорему 2.

*Следствие*. Точечная оценка (7) для показателя $\overline{\Upsilon}(t)$ несмещенная.

В самом деле, согласно (7), имеем

$$\hat{\hat{\Upsilon}}_n(t)=1-\hat{\Upsilon}_n(t),$$

откуда, с учетом (9), находим

$$M\left(\hat{\overline{\Upsilon}}(t)\right)=\overline{\Upsilon}(t),$$

где показатель $\overline{\Upsilon}(t)$ определен соотношением (2).

## 2. Нижняя доверительная граница показателя СДБН

При малых объемах выборки $n$ степень доверия к точечной оценке показателя СДБН $\hat{\Upsilon}_n(t)$ крайне низка. Поэтому докажем следующее утверждение:

*Теорема 3*. Пусть $p$ — заданная доверительная вероятность. Тогда нижней доверительной границей показателя СДБН служит следующая величина:

$$\underline{\Upsilon}_n(t)=\hat{\Upsilon}_n(t)-\sqrt{\frac{-\ln(1-p)}{2n}}. \qquad (13)$$

*Доказательство*. Для установления формулы (13) воспользуемся неравенством Хевдинга [4]:

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \ge \varepsilon\right) \le \exp\left(-\frac{2n^2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)}\right), \quad (14)$$

где $\varepsilon > 0$ — произвольное число; $X_i$ — случайная величина, удовлетворяющая условиям $a_i \le X_i \le b_i$, $(i = 1, 2, ..., n)$; $\mu = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$.

Примем следующие обозначения:

$$X_i = \begin{cases} \tau_i / t, \text{если} \quad \tau_i \in (0, t); \\ 1, \quad \text{если} \quad \tau_i \ge t; \end{cases}$$

$$a_i = 0,\ b_i = 1 \text{ при } i = 1, 2, ..., n.$$

Тогда согласно (6) имеем

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\Upsilon}_n(t). \quad (15)$$

Для определения величины $\mu$ воспользуемся теоремой 2, согласно которой

$$\mu = \Upsilon(t). \quad (16)$$

Далее, используя соотношения (15) и (16) в неравенстве (14), находим

$$\Pr\left(\widehat{\Upsilon_n}(t) - \Upsilon(t) \ge \varepsilon\right) \le \exp\left(-2n\varepsilon^2\right),$$

т. к. $\sum_{i=1}^{n}(b_i - a_i) = n$.

Следовательно, имеем

$$\Pr\left(\hat{\Upsilon}_n(t) - \Upsilon(t) < \varepsilon\right) > 1 - \exp\left(-2n\varepsilon^2\right),$$

откуда находим

$$\Pr\left(\Upsilon(t) > \hat{\Upsilon}_n(t) - \varepsilon\right) > 1 - \exp\left(-2n\varepsilon^2\right). \quad (17)$$

Поскольку число $\varepsilon > 0$ — произвольное, то выберем его из следующего условия:

$$1 - \exp\left(-2n\varepsilon^2\right) = p, \quad (18)$$

где $p$ — заданная доверительная вероятность.

Решая уравнение (18), находим

$$\varepsilon = \sqrt{\frac{-\ln(1-p)}{2n}}.$$

Следовательно, согласно соотношениям (17) и (18), имеем

$$\Pr\left(\Upsilon(t) > \underline{\Upsilon}_n(t)\right) > p,$$

где значение $\underline{\Upsilon}_n(t)$ определено соотношением (13), что и доказывает теорему 3.

## 3. Определение минимального количества объектов, необходимого для проведения выборочного эксперимента

Формула (13) позволяет определить минимальное количество объектов, необходимое для проведения выборочного эксперимента. Для этой цели докажем следующее утверждение:

*Теорема 4.* Пусть $p$ — заданная доверительная вероятность $(0 < p < 1)$; $\underline{\Upsilon}_n(t)$ — нижняя доверительная граница показателя $\Upsilon(t)$. Тогда минимальное количество наблюдаемых объектов в течение времени $t$, необходимое для проведения выборочного эксперимента, определяется по формуле:

$$n = \begin{cases} \dfrac{-\ln(1-p)}{2\left(1-\underline{\Upsilon}_n(t)\right)^2}, \\ \qquad \text{если правая часть — целое число;} \\ \left[\dfrac{-\ln(1-p)}{2\left(1-\underline{\Upsilon}_n(t)\right)^2}\right] + 1, \quad \text{иначе,} \end{cases} \quad (19)$$

где $[\cdot]$ — целая часть выражения, стоящего внутри скобок.

*Доказательство.* Воспользуемся формулой (13), из которой найдем

$$n = \frac{-\ln(1-p)}{2\left(\hat{\Upsilon}_n(t) - \underline{\Upsilon}_n(t)\right)^2}. \quad (20)$$

Поскольку в силу свойства 1 для показателя СДБН имеем $\hat{\Upsilon}_n(t) \le 1$, то, учитывая это в соотношении (20), получим

$$n \ge \frac{-\ln(1-p)}{2\left(1-\underline{\Upsilon}_n(t)\right)^2},$$

откуда найдем искомое соотношение (19).

Из доказанной формулы (19) следует, что если доверительная вероятность $p$ стремится к 1, то объем выборки увеличивается и, напротив, если $p$ уменьшается $(p \to 0)$, то количество объектов для проведения выборочного эксперимента также уменьшается.

Проводя сравнительный анализ влияния принимаемых значений $\underline{Y}_n(t)$ на величину объема выборки, делаем следующий вывод: если значение нижней доверительной границы, равное $\underline{Y}_n(t)$, стремится к 1, то объем выборки увеличивается и, напротив, при уменьшении величины $\underline{Y}_n(t)$ объем выборки уменьшается.

Оба вывода хорошо согласуются с логикой проведения выборочного эксперимента.

*Пример.* Определить минимальное количество однотипных объектов для проведения выборочного эксперимента при условии, что с доверительной вероятностью, равной 0,865, нижняя доверительная граница показателя $Y(t)$ не менее $2/3$.

*Решение.* Согласно условиям примера имеем $p = 0{,}865$; $\underline{Y}_n(t) \geq 2/3$. Тогда

$$\frac{-\ln\left(1-0{,}865\right)}{2\left(1-\underline{Y}_n(t)\right)^2} \geq 9,$$

откуда, согласно (19), находим, что минимальное количество однотипных объектов для проведения выборочного эксперимента равно 9.

## Заключение

В данной работе на основе сравнения показателей надежности отказавших и неотказавших однотипных технических объектов в течение заданного времени определяется объем выборки — минимальное количество объектов для проведения выборочного эксперимента. Полученная расчетная формула для объема выборки может быть использована не только при планировании испытаний объектов на надежность, но и при планировании и проведении других видов экспериментов в различных областях науки.

## Литература


1. *Сидняев Н. И.* Теория планирования эксперимента и анализ статистических данных. М.: Юрайт, 2011. 399 с.
2. *Давыдов А. А.* Системный подход в социологии: новые направления, теории и методы анализа социальных систем. М.: URSS, 2005. 396 с.
3. *Кэмпбелл Д.* Модели экспериментов в социальной психологии и прикладных исследованиях. СПб.: Изд-во «Социально-психологический центр», 1996. 391 с.
4. *Hoeffding W.* Probability inequalities for sums of bounded random variables // Annal. Stat. Assoc. 58, 1963. P. 13–29.



**Садыхова Лала Гуламовна**. Доцент МГТУ им. Н. Э. Баумана. К. к. н. Окончила МГУ в 1998 г. Количество печатных работ: 46. Область научных интересов: теория межкультурной коммуникации, модели коммуникации, методы исследования социокультурных процессов. E-mail: lala_sad@inbox.ru

**Назаренко Дмитрий Борисович**. Аспирант МГТУ им. Н. Э. Баумана. Окончил МГТУ им. Н. Э. Баумана в 1989 г. Количество печатных работ: 3. Область научных интересов: теория надежности, планирование эксперимента. E-mail: gendir@odinled.ru

**Бабаев Ислам Акмурадович**. Студент МГТУ им. Н. Э. Баумана. Количество печатных работ: 2. Область научных интересов: математические методы и модели планирования выборочного эксперимента. E-mail: sleavik@gmail.com