Министерство образования и науки Российской Федерации

Нижегородский государственный университет им. Н.И. Лобачевского

Национальный исследовательский университет

М.С. Тихов
Т.С. Бородина

Эконометрические модели с цензурированными данными

Учебно-методическое пособие

Нижний Новгород
2012

УДК 519.2
ББК В171(Я73-4)
Т 46

Т 46 Тихов М.С., Бородина Т.С. ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ С ЦЕНЗУРИРОВАННЫМИ ДАННЫМИ: Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2012. – 50 с.

Рецензент: доц. кафедры МФ, к.ф-м.н. **А.В. Калинин**

Рассматриваются вопросы эконометрического анализа по случайно цензурированным выборкам. Изучаемая тема входит как составляющая часть в общий курс «Многомерные статистические методы» для студентов механико-математического факультета, обучающихся по направлению 080500 «Бизнес-информатика», а также общего курса «Современные проблемы прикладной теории вероятностей» для магистров факультета ВМК, обучающихся по направлению 010500 «Прикладная математика и информатика». Данное пособие рассматривает проблему эконометрического анализа по случайно цензурированным выборкам и состоит из теоретической части, связанной с построением оценок Каплана-Мейера и Нельсона-Аалена и их характеристик, а также из практической части, связанной с анализом срока службы и выбытия финансовых активов. Кроме того рассмотрена тема определения остаточного срока службы оборудования и машин, связанная с оценкой оставшегося ресурса оборудования и их остаточной стоимости.

Пособие предназначено для студентов 3 курса механико-математического факультета, а также для магистров 2 года обучения факультета ВМК.

УДК 519.2
ББК В171(Я73-4)

Введение

Мы будем рассматривать неполные выборки, которые довольно часто встречаются в экономических исследованиях. Например, если мы захотим оценивать среднюю продолжительность жизни населения, то можно отметить следующее: у нас имеются как полные выборки, то есть мы имеем для некоторых индивидуумов время начало жизни и время конца жизни – точную продолжительность их жизни (будем употреблять термин *отказ*), а для некоторых только время начало жизни (будем называть это *приостановкой*), и эти индивидуумы могут оказывать сильное влияние на среднюю продолжительность жизни. Представим себе гипотетический случай, когда в нашем населенном пункте имеется высокая детская смертность и есть много глубоких стариков, которых мы не можем не учитывать (исходя из полученных цифр, мы можем планировать пенсионный фонд и т.д.). В таком случае большое количество неотказавших элементов выборки могут оказывать сильное влияние на оценку математического ожидания. Классическая теория вероятностей и математическая статистика предполагает наличие только полной выборки и не имеет методов оценки таких выборок, которые появились в последнее время.

Кроме того, важным вопросом является также оценка остаточного срока службы нематериальных активов, оценка среднего времени их жизни, важным вопросом является также оценка остаточного срока службы машин. В эту тему входят и модели бинарного выбора.

1. Модели бинарного выбора.

Метод наименьших квадратов основан на предположении, что зависимые переменные y являются непрерывными. В некоторых случаях это предположение не выполняется (они являются дискретными), и это вносит определенные трудности в исследование регрессионных моделей с такими показателями.

Рассмотрим следующий пример.

Пример 1. Факторы некредитоспособности российских банков. Речь идет о том, как те или иные характеристики влияют на жизнеспособность банков? Для исследования этой проблемы будем рассматривать бинарную переменную, принимающую значение 1 или 0, в зависимости от того, находится ли банк в критическом состоянии или нет. Решение о том, является ли банк проблемным или нет, принималось на основании рейтинга банков, опубликованного в журнале «Профиль» от 21 июня 1999 г. Значение 1 приписывалось банкам с отрицательным капиталом; банкам, имеющим 4-ю группу проблемности; банкам, у которых отозвана лицензия или принято решение об отзыве. Остальным банкам присвоено значение 0. Таким образом, зависимая переменная принимает два значения: 1 и 0. Что касается независимых переменных, то из многочисленных характеристик банков в модель включены следующие характеристики:

x_1: СУМОБ – суммарные обязательства (тыс. руб.);

x_2: ВАЛС – валютная составляющая (%);

x_3: НЕДВ/ЧАКТ, где НЕДВ – недвижимость (тыс. руб.), ЧАКТ – чистые активы (тыс. руб.);

x_4: ПРИБ/ЧАКТ, ПРИБ – прибыль (убыток) (тыс. руб.);

x_5: СЧЛ/СУМОБ, где СЧЛ – средства частных лиц (тыс. руб.);

x_6: СУМОБ/РИСКАКТ, где РИСКАКТ – работающие рисковые активы (тыс. руб.).

Набор характеристик представим в виде вектора $\mathbf{x}' = (x_1, x_2, \dots, x_6)$ – независимые переменные. Тогда модель имеет вид: $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, $i = 1, \dots, n$.

Задача – оценить неизвестные параметры $\boldsymbol{\beta}$.

Линейная модель вероятности.

Рассмотрим сначала обычную линейную модель регрессии:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t\,, \quad t = 1,\ldots,n. \text{ т.е. } \mathbf{E}(y_t) = \boldsymbol{x}_t'\boldsymbol{\beta}\ . \qquad (1)$$

где t – номер наблюдения, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_k)'$ – набор неизвестных параметров (коэффициентов), ε_t – случайная ошибка (непрерывная случайная величина с нулевым математическим ожиданием). Найдем математическое ожидание величины y_t. Так как y_t принимает значения 0 или 1 и $\mathbf{E}(\varepsilon_t) = 0$, то

$$\mathbf{E}(y_t) = 1 \cdot \mathbf{P}(y_t = 1) + 0 \cdot \mathbf{P}(y_t = 0) = \mathbf{P}(y_t = 1) = \boldsymbol{x}_t'\boldsymbol{\beta}\,.$$

Таким образом, модель может быть записана в виде

$$\mathbf{P}(y_t = 1) = \boldsymbol{x}_t'\boldsymbol{\beta}\,, \qquad (2)$$

поэтому ее называют *линейной моделью вероятности.*

Отметим ***следующее***:

1) из соотношения (1) следует, что ошибка ε в каждом наблюдении может принимать только два значения: $\varepsilon_t = 1 - \boldsymbol{x}_t'\boldsymbol{\beta}$ с вероятностью $\mathbf{P}(y_t = 1)$ и $\varepsilon_t = -\boldsymbol{x}_t'\boldsymbol{\beta}$ с вероятностью $\mathbf{P}(y_t = 0)$, что противоречит непрерывности величин ε_t;

2) дисперсия ошибки ε_t будет равна $\mathbf{D}(\varepsilon_t) = \boldsymbol{x}_t'\boldsymbol{\beta}(1 - \boldsymbol{x}_t'\boldsymbol{\beta})$, т.е. зависит от $\boldsymbol{x}_t$ (модель гетероскедастична), поэтому мы не можем воспользоваться обычным методом наименьших квадратов;

3) прогнозные значения $\hat{y}_t = \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}$, которые по смыслу являются вероятностью могут не лежать в интервале $[0,1]$ ($\hat{\boldsymbol{\beta}}$ – оценка коэффициентов $\boldsymbol{\beta}$), поэтому мы рассмотрим несколько иной подход.

Probit- и *logit*-модели.

Основной недостаток линейной модели вероятности есть следствие предположения о линейной зависимости вероятности $\mathbf{P}(y_t = 1)$ от $\boldsymbol{\beta}$. Его можно преодолеть, если считать, что

$$\mathbf{P}(y_t = 1) = F(\boldsymbol{x}_t'\boldsymbol{\beta})\,, \qquad (3)$$

где $F(x)$ – некоторая функция, область значений которой лежит в интервале $[0,1]$, в частности в качестве $F(x)$ можно взять функцию распределения некоторой случайной величины. Наиболее часто в качестве функции $F(x)$ используют:

– функцию стандартного нормального распределения

$$F(u) = \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{x^2}{2}}\,dx\,, -\infty < u < +\infty,$$

(соответствующую модель называют *probit-моделью*);

– функцию логистического распределения

$$F(u) = \Lambda(u) = \frac{e^u}{1+e^u}, -\infty < u < +\infty,$$

(соответствующую модель называют *logit-моделью*).

Графики функций $\Phi(u)$ и $\Lambda(u)$ приведены на рис. 1 и рис. 2 соответственно:

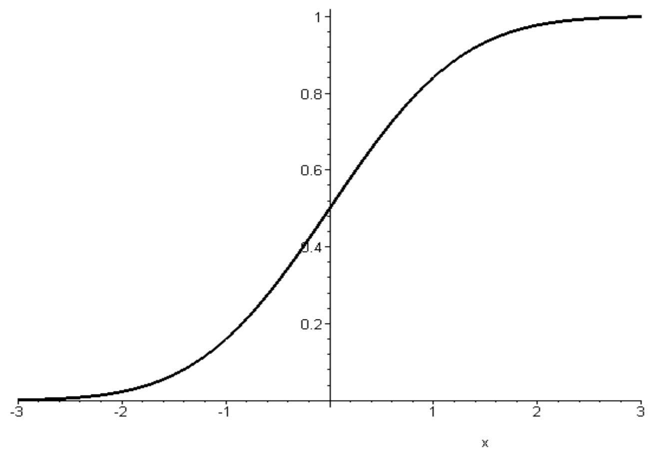


Рис. 1. График функции стандартного нормального распределения $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$.

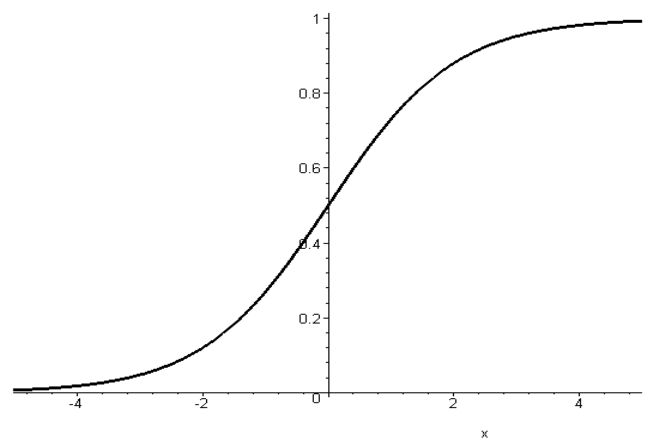


Рис. 2. График функции логистического распределения $\Lambda(x) = \frac{e^x}{1+e^x}$.

Рассмотрим следующую интерпретацию модели (3). Пусть

$$y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t,$$

где ошибки ε_t независимы и одинаково распределены с нулевым средним, дисперсией σ^2 и функцией распределения нормированной случайной ошибки ε_t/σ, равной $F(x)$. Решение $y_t = 1$ принимается тогда, когда y_t^* превосходит некоторое неслучайное пороговое значение. Без ограничения общности, если константа включена в число регрессоров, можно считать это пороговое значение равным нулю. Таким образом, $y_t = I(y_t^* \geq 0)$.

Предположим, что случайные ошибки имеют одно и то же симметричное распределение $F(x) = 1 - F(-x)$. Тогда

$$\mathbf{P}(y_t = 1) = \mathbf{P}(y_t^* \geq 0) = \mathbf{P}(\boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \geq 0) = \mathbf{P}(\varepsilon_t \geq -\boldsymbol{x}_t'\boldsymbol{\beta}) = \mathbf{P}(\varepsilon_t \leq \boldsymbol{x}_t'\boldsymbol{\beta}) = F\left(\frac{\boldsymbol{x}_t'\boldsymbol{\beta}}{\sigma}\right), \quad (4)$$

что с точностью до нормировки совпадает с (3). Поскольку в модели (4) параметры $\boldsymbol{\beta}$ и σ участвуют в виде отношения, то можно оценить только отношение $\frac{\boldsymbol{\beta}}{\sigma}$, поэтому будем считать, что $\sigma = 1$.

Вопрос о том, какую из моделей (*probit* или *logit*) использовать является предметом применения модели для конкретной ситуации. График функций $\Phi(x)$ и $\Lambda(x)$ примерно одинаков (см. рис. 1 и 2), в то же время «хвосты» логистического распределения тяжелее «хвостов» нормального распределения. Так, например, $\Phi(x) \sim 1 - \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ при $x \to \infty$, а для логистического распределения $\Lambda(x) \sim 1 - e^{-x}$, $x \to \infty$. Здесь можно дать следующую рекомендацию: для выборок с небольшим объемом выводы, получаемые с помощью *probit*- и *logit*-модели, будут примерно одинаковыми. Кроме того, $y = \ln \frac{\Lambda(\boldsymbol{x}_t'\boldsymbol{\beta})}{1 - \Lambda(\boldsymbol{x}_t'\boldsymbol{\beta})} = \boldsymbol{x}_t'\boldsymbol{\beta}$, т.е. имеем обычную схему МНК.

Пример 2. Пусть в выборке $y = 1$, если количество занятых в фирме выросло ($y = 0$ – в противном случае); x_1 – доход фирмы в млн. долл.; $x_2 = 1$, если фирма относится к области высоких технологий ($x_2 = 0$ – в противном случае). Получена следующая модель:

$$\hat{z} = 0.40 + 0.20 \cdot x_1 + 0.10 \cdot x_2.$$

Требуется определить оценку вероятности роста занятости для высокотехнологичной фирмы А с доходом в 5 млн. долл., и для фирмы Б, не относящейся к сфере высоких технологий и имеющей доход в 7 млн. долл.

Решение. Рассчитаем сначала значения функции $\hat{z}$:

$$\hat{z}_1 = 0.4 + 0.2 \cdot 5 + 0.1 \cdot 1 = 1.5,$$
$$\hat{z}_2 = 0.4 + 0.2 \cdot 7 + 0.1 \cdot 0 = 1.8.$$

Для *logit*-модели вероятность роста занятости определяется следующим образом:

$$\mathbf{P}(y_t = 1) = \frac{1}{1 + e^{-z_t}}.$$

Соответственно, получим:

$$\mathbf{P}(y_t = 1) = \frac{1}{1+e^{-1.5}} = 0.8176 \qquad \text{(для фирмы А)},$$

$$\mathbf{P}(y_t = 1) = \frac{1}{1+e^{-1.8}} = 0.8581 \qquad \text{(для фирмы Б)}.$$

Таким образом, вероятности роста занятости в фирмах А и Б составят соответственно 81,76% и 85,81%.

Для *probit*-модели вероятность роста занятости определяется следующим образом:

$$\mathbf{P}(y_t = 1) = \Phi(z_t).$$

Тогда

$$\mathbf{P}(y_t = 1) = \Phi(1.5) = 0.9332 \qquad \text{(для фирмы А)},$$

$$\mathbf{P}(y_t = 1) = \Phi(1.8) = 0.964 \qquad \text{(для фирмы Б)}.$$

Оценивание модели.

Для оценивания параметров $\boldsymbol{\beta}$ модели (3) обычно используют метод максимального правдоподобия. Предположим, что наблюдения $y_1, y_2, \ldots, y_n$ независимы. Функция правдоподобия равна

$$L = L(y_1, \ldots, y_n) = \prod_t (F(\boldsymbol{x}_t'\boldsymbol{\beta}))^{y_t} (1 - F(\boldsymbol{x}_t'\boldsymbol{\beta}))^{1-y_t} .$$

Логарифмируя, получаем:

$$l = \ln L = \sum_t (y_t \ln F(\boldsymbol{x}_t'\boldsymbol{\beta}) + (1-y_t) \ln (1 - F(\boldsymbol{x}_t'\boldsymbol{\beta}))). \qquad (5)$$

Дифференцируя равенство (5) по $\boldsymbol{\beta}$, приходим к векторному уравнению

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_t \left(\frac{y_t\, f(\boldsymbol{x}_t'\boldsymbol{\beta})}{F(\boldsymbol{x}_t'\boldsymbol{\beta})} - \frac{(1-y_t)\, f(\boldsymbol{x}_t'\boldsymbol{\beta})}{1 - F(\boldsymbol{x}_t'\boldsymbol{\beta})} \right) \boldsymbol{x}_t = \mathbf{0}. \qquad (6)$$

Для *logit*-модели это уравнение упрощается. Пользуясь тождеством $\Lambda'(x) = \Lambda(x)(1 - \Lambda(x))$, имеем

$$\sum_t (y_t - \Lambda(\boldsymbol{x}_t'\boldsymbol{\beta}))\, \boldsymbol{x}_t = \mathbf{0}. \qquad (7)$$

Уравнение правдоподобия (6) есть лишь необходимое условие локального экстремума. Покажем, что для *logit*-модели (а также для *probit*-модели) логарифмическая функция правдоподобия l является вогнутой функцией по $\boldsymbol{\beta}$ и, значит, решение уравнения (6) доставляет максимум функции l. Действительно,

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}'} = \sum_t (y_t - \Lambda(\boldsymbol{x}_t'\boldsymbol{\beta}))\, \boldsymbol{x}_t',$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'} = -\sum_t \Lambda^2(\boldsymbol{x}_t'\boldsymbol{\beta})(1 - \Lambda(\boldsymbol{x}_t'\boldsymbol{\beta}))\, \boldsymbol{x}_t \boldsymbol{x}_t'.$$

Последняя матрица является суммой неотрицательно определенных матриц $\boldsymbol{x}_t\boldsymbol{x}_t'$ с отрицательными коэффициентами, поэтому является неположительно определенной матрицей

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'} \le \mathbf{0},$$

что и требовалось доказать.

Пример 3. Рассмотрим модель бинарного выбора $\mathbf{P}(y_t = 1) = F(\alpha + \beta d_t)$, где d – фиктивная переменная (принимающая значения 0 или 1). Ниже представлены результаты 20 наблюдений:

		y 0	1	
d	0	4	6	10
	1	8	2	10
		12	8	20

а) Оцените параметры α, β, используя *logit*-модель. Проверьте гипотезу $\mathbf{H_0}: \beta = 0$

Решение. Рассмотрим модель $\mathbf{P}(y_t = 1) = F(\alpha + \beta d_t)$. Функция правдоподобия равна

$$L = F^6(\alpha) F^2(\alpha+\beta)(1 - F(\alpha))^4 (1 - F(\alpha+\beta))^8.$$

Обозначим $\gamma = \alpha + \beta$. Тогда

$$\ln L = 6 \ln F(\alpha) + 2 \ln F(\gamma) + 4 \ln(1 - F(\alpha)) + 8 \ln(1 - F(\gamma)).$$

Оценка максимального правдоподобия является решением системы уравнений:

$$\begin{cases} \dfrac{\partial \ln L}{\partial \alpha} = 6\, \dfrac{f(\alpha)}{F(\alpha)} - 4 \dfrac{f(\alpha)}{1 - F(\alpha)} = 0, \\ \dfrac{\partial \ln L}{\partial \gamma} = 2\, \dfrac{f(\gamma)}{F(\gamma)} - 8 \dfrac{f(\gamma)}{1 - F(\gamma)} = 0. \end{cases}$$

Для *logit*-модели $F(x) = \Lambda(x)$, $\dfrac{\Lambda'(x)}{\Lambda(x)} = 1 - \Lambda(x)$, $\dfrac{\Lambda'(x)}{1 - \Lambda(x)} = \Lambda(x)$, получаем:

$$\begin{cases} 6\,(1 - \Lambda(\alpha)) - 4\Lambda(\alpha) = 0, \\ 2\,(1 - \Lambda(\gamma)) - 8\Lambda(\gamma) = 0. \end{cases}$$

Решением является

$$\Lambda(\alpha) = 0.6,\ \Lambda(\gamma) = 0.2, \text{ откуда } \frac{1}{1 + e^{-\alpha}} = 0.6,\ \frac{1}{1 + e^{-\gamma}} = 0.2, \text{ или}$$

$$\alpha = \ln 1.5 = 0.405,\ \gamma = -\ln 4 = -1.386,\ \beta = -1.791.$$

Для проверки гипотезы $\beta = 0$ применим критерий отношения правдоподобия в виде: $LR = -2(\ln \tilde{L} - \ln \hat{L})$ (при гипотезе $H_0: \beta = 0$ LR имеет при $n \to \infty$ $\chi^2(1)$ – распределение, поэтому гипотеза отвергается, если $LR > \chi^2_{1-\alpha}(1)$).

Значение функции правдоподобия для регрессии без ограничения равно

$$\ln \hat{L} = 6 \ln 0.6 + 2 \ln 0.2 + 4 \ln 0.4 + 8 \ln 0.8 = -11.73.$$

Оценим регрессию с ограничением $\beta = 0$:

$$L = F^8(\alpha)(1-F(\alpha))^{12}, \qquad \ln L = 8 \ln F(\alpha) + 12 \ln(1-F(\alpha)),$$

$$\frac{\partial \ln L}{\partial \alpha} = 8 \frac{f(\alpha)}{F(\alpha)} - 12 \frac{f(\alpha)}{1-F(\alpha)} = 0, \quad 8(1-\Lambda(\alpha)) - 12\Lambda(\alpha) = 0 \;\Rightarrow\; 0.4 = \frac{1}{1+e^{-\alpha}} \;\Rightarrow\; \alpha = \ln\frac{2}{3} = -0.405,$$

$$\ln \tilde{L} = 8 \ln 0.4 + 12 \ln 0.6 = -13.46.$$

Получаем, что $LR = 3.46$, т.е. гипотеза $\beta = 0$ не отвергается при уровне значимости 0.05, поскольку критическое значение равно $\chi^2_{0.95} = 3.84$.

На самом деле, конечно, функция распределения $F(x)$ неизвестна, и для ее оценки можно использовать ядерные оценки:

$$\hat{F}(x) = \frac{1}{nh}\sum_{i=1}^{n} y_i K\left(\frac{x - \boldsymbol{x}_i^T\beta}{h}\right) \Bigg/ \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - \boldsymbol{x}_i^T\beta}{h}\right),$$

где в качестве ядерной функции $K(x)$ можно взять $K_0(x) = 0,75(1-x^2)I(|x| \le 1)$, $h = h(n)$– ширина окна просмотра данных. Обычно в качестве h берут $cn^{-1/5}$, константа c определяется из так называемой процедуры кросс-валидейшн.

Клейн и Спэйди предложили оценивать неизвестные параметры β методом максимального правдоподобия. Оценка логарифмической функции правдоподобия имеет вид

$$L(\beta, h) = \sum_i (1-y_i)\ln(1-F_{-i}(\boldsymbol{x}_i^T\beta)) + \sum_i y_i \ln F_{-i}(\boldsymbol{x}_i^T\beta),$$

где $F_{-i}(\boldsymbol{x}_i^T\beta)$ есть оценка функции распределения по всей выборке за исключением i-го наблюдения. Максимизация по β и h дает полупараметрическую ММП-оценку параметра β.

2. Цензурированные выборки. Сокращение времени испытания за счет цензурирования.

При испытаниях изделий на продолжительность эксплуатации в случаях высоконадежных изделий такие испытания могут продолжаться достаточно долго. В теории надежности для сокращения времени испытаний иногда применяют метод форсированных испытаний. Но в этом случае встает проблема пересчета времени форсированных испытаний во времена испытаний в нормальных условиях. Другой способ, который можно применить – это использование цензурированных выборок. Рассмотрим случай, когда испытания независимы, а закон распределения наработки до отказа i – го изделия показательный с плотностью распределения

$$f_i(x) = \exp(-\lambda_i x)\lambda_i\,,\; x > 0;\;\; \lambda_i = 1/(\beta_i\theta);\theta\,,\beta_i > 0\,,1 \le i \le n\,, \qquad (8)$$

где θ – неизвестный параметр, а β_i – известные коэффициенты. Параметры β_i могут трактоваться как коэффициенты форсирования испытаний по сравнению с обычными условиями эксплуатации. Задача состоит в оценке параметра θ по наблюдениям за отказом k первых отказавших изделий, поставленных на испытания.

В случае, когда $x_1, x_2, \ldots, x_n$ – независимые и одинаково распределенные по показательному закону случайные величины ($\beta_i = 1, i = 1, 2, \ldots, n$), наилучшей в определенном смысле оценкой для параметра θ является статистика

$$t_n^{(k)} = \frac{1}{k}\cdot\left(\sum_{j=1}^{k} x_n^{(j)} + (n-k)\,x_n^{(k)}\right),$$

где $x_n^{(1)} < x_n^{(2)} < \ldots < x_n^{(n)}$ – вариационный ряд, построенный по $x_1, x_2, \ldots, x_n$. Она обладает следующими свойствами: $t_n^{(k)}$ есть несмещенная оценка параметра θ с дисперсией $\mathbf{D}_\theta(t_n^{(k)}) = \theta^2/k$, не зависящей от n. Последнее обстоятельство дает возможность сравнить ее с оценкой $t_k^{(k)}$ по полной выборке объема k, дисперсия которой равна $\mathbf{D}_\theta(t_k^{(k)}) = \theta^2/k$, т.е. $\mathbf{D}_\theta(t_n^{(k)}) = \mathbf{D}_\theta(t_k^{(k)})$, но средняя длительность испытания для оценки $t_n^{(k)}$ равна

$$\tau_1 = \mathbf{E}_\theta(X_n^{(k)}) = \theta\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n-k+1}\right),$$

в то время как для оценки $t_k^{(k)}$ среднее время испытания равно

$$\tau_2 = \mathbf{E}_\theta(X_k^{(k)}) = \theta\left(\frac{1}{k} + \frac{1}{k-1} + \ldots + \frac{1}{2} + 1\right).$$

Следовательно, если $n \ge k$, то $\tau_1 \le \tau_2$. Так при $n = 10, k = 5,\; \tau_1/\tau_2 = 0.283$. Эти выводы легко получить используя представление Реньи для показательного распределения, которое имеет вид

$$x_n^{(k)} = \frac{\theta\xi_1}{n} + \frac{\theta\xi_2}{n-1} + \ldots + \frac{\theta\xi_k}{n-k+1}\,,$$

$1 \le k \le n$,где $\xi_1, \xi_2, \dots, \xi_n$ – независимые одинаково распределенные с.в. с функцией распределения $F(x) = \max(0, 1 - \exp(-x))$. Это представление было обобщено в 1991г. автором (Тихов М.С. О сокращении длительности испытаний при цензурировании выборки. – ж. «*Теория вероятностей и её применения*», 1991, т.36, в.3, с.626-629.)

Определим сначала понятие *ранга* и *антиранга*. Для вектора $x = (x_1, x_2, \dots, x_n)$, у которого никакие две координаты не совпадают, пусть $r_i(x)$ – число координат, не превосходящих x_i, т.е. $x_i = x_n^{(r_i)}$. Статистика $R_i = r_i(X)$ будет называться *рангом* X_i.

Пусть $(d_1, d_2, \dots, d_n)$ – обратная перестановка по отношению к $(r_1, r_2, \dots, r_n)$, т.е.

$$r_{d_i} = d_{r_i} = i, \ 1 \le i \le n,$$

и пусть $D = (D_1, D_2, \dots, D_n)$ обратная в этом смысле перестановка к $R = (R_1, R_2, \dots, R_n)$. Статистики $D_1, D_2, \dots, D_n$ будем называть *антирангами*.

Основной результат (обобщение представления Реньи) состоит в следующем. Если $X = (X_1, X_2, \dots, X_n)$ имеют распределение (8), то

$$X_n^{(k)} \overset{d}{=} \frac{\xi_1}{\lambda_{D_1} + \lambda_{D_2} + \dots + \lambda_{D_n}} + \frac{\xi_2}{\lambda_{D_2} + \dots + \lambda_{D_n}} + \dots + \frac{\xi_k}{\lambda_{D_k} + \dots + \lambda_{D_n}} .$$

В том случае, когда распределения одинаковые ($\lambda_1 = \lambda_2 = \dots = \lambda_n = \lambda$), то

$$X_n^{(k)} = \frac{\xi_1}{n\lambda} + \frac{\xi_2}{(n-1)\lambda} + \dots + \frac{\xi_k}{(n-k+1)},$$

т.е. имеем представление Реньи. Так как $\mathbf{E}(\xi_i) = 1$ и ожидание суммы равно сумме ожиданий, то имеем выражения для количеств τ_1 и τ_2.

3. Случайно цензурированные выборки. Оценки Каплана-Мейера и оценки Нельсона-Аалена.

Классическая математическая статистика имеет дело со следующей ситуацией. Пусть $X_1, X_2, \dots, X_n$ – независимые и одинаково распределенные случайные величины с неизвестной функцией распределения $F(x)$. Рассмотрим эмпирическую функцию распределения $\hat{F}_n(x)$, где

$$\hat{F}_n(x) = \frac{\sum_{i=1}^{n} I(X_i < x)}{n},$$

где $I(A)$ – индикатор события A, т.е. $I(A)=1$, если событие A произошло, и равно нулю в противном случае. При увеличении объема выборки, т.е. при $n \to \infty$ имеем:

1) $\hat{F}_n(x) \underset{n\to\infty}{\overset{p}{\to}} F(x)$ (теорема Бернулли) для каждого фиксированного x;

2) $\mathbf{P}(\hat{F}_n(x) \underset{n\to\infty}{\to} F(x)) = 1$ (теорема Бореля) для каждого фиксированного x;

3) Пусть $\Delta_n = \sup\limits_x |\hat{F}_n(x) - F(x)|$. Тогда

$$\mathbf{P}(\Delta_n \underset{n\to\infty}{\to} 0) = 1 \quad \text{(теорема Гливенко) – равномерная сходимость.}$$

Результаты 1) и 2) говорят о точечной сходимости, утверждение 3) – о равномерной сходимости. Более того,

4) $\sqrt{n}\,(\hat{F}_n(x) - F(x)) \underset{n\to\infty}{\overset{d}{\to}} \xi \in N(0, F(x)(1-F(x)))$ (теорема Муавра-Лапласа – точечная сходимость);

5) $\mathbf{P}(\sqrt{n}\cdot\Delta_n < x) \underset{n\to\infty}{\to} K(x) = \sum\limits_{j=-\infty}^{+\infty} (-1)^j \exp(-2j^2x^2)$, $x > 0$, (теорема Колмогорова).

Однако, практически во всех экспериментах по наблюдению за объектами, связанными со временем жизни, присутствуют неполные данные. Это связано с тем, что некоторые объекты могут не наблюдаться в течение полного времени до отказа. Мы рассмотрим только лишь, так называемые случайно цензурированные выборки, которые опишем следующим образом.

Пусть $\{(X_i, Y_i), 1 \le i \le n\}$ – последовательность независимых и одинаково распределенных случайных векторов, причем X_j и Y_j независимы также между собой. Обозначим через $F(x)$ и $G(x)$ их функции распределения соответственно.

Введем величины $\varsigma_j = \min(X_j, Y_j)$ и $w_j = I(X_j < Y_j)$. Тогда совокупность $\{(\varsigma_j, w_j), 1 \le j \le n\}$ будем называть ***случайно цензурированной выборкой*** (СЦВ). Таким образом, мы наблюдаем отказ, если $I(X_j < Y_j) = 1$ в момент x_j, или приостановку, если $I(X_j < Y_j) = 0$ в момент y_j, т.е. мы знаем момент ς_j и, кроме того, мы знаем, что мы наблюдаем – отказ или приостановку. Такие выборки называют также право цензурированными выборками.

Если $\mathbf{P}(Y=c)=1$, то мы имеем цензурированные выборки I типа, которые будем рассматривать как частный случай СЦВ.

Нас будет интересовать оценка функции распределения $F(x)$ по СЦВ.

Пусть $S(x)=1-F(x)$. Для этих целей обычно используется оценка **Каплана-Мейера**, определяемая следующим соотношением:

$$\hat{S}_n(x)=\prod_{j:\varsigma_j<x}\left(\frac{n-j}{n-j+1}\right)^{w_j},$$

где мы будем считать $\{\varsigma_j\}$ упорядоченными.

Известно, что $\hat{S}_n(x)\xrightarrow[n\to\infty]{p}S(x)$ и $\sqrt{n}(\hat{S}_n(x)-S(x))$ асимптотически нормальна (при фиксированном x).

Попробуем понять, как получается оценка Каплана-Мейера в указном виде. Пусть $S(x)=1-F(x)=\mathbf{P}(X\geq x)$ и $f(x)=F'(x)-$ плотность распределения. Тогда

$$f(t)=\lim_{\delta\to0+}\frac{\mathbf{P}(t\leq X<t+\delta)}{\delta}.$$

Введем интенсивность отказа

$$h(t)=\lim_{\delta\to0+}\frac{\mathbf{P}(t\leq X<t+\delta|t\leq X)}{\delta}\approx\frac{\mathbf{P}(t\leq X<t+\delta|t\leq X)}{\delta}\ -$$

вероятность отказа при условии, что $X\geq t$, т.е. отказ произойдет позже момента t. Имеем

$$\mathbf{P}(t\leq X<t+\delta|t\leq X)=\frac{\mathbf{P}((t\leq X<t+\delta)(\,t\leq X))}{\mathbf{P}(\,t\leq X)}=\frac{\mathbf{P}(t\leq X<t+\delta)}{\mathbf{P}(\,t\leq X)}\approx\frac{f(t)\delta}{1-F(t)},\ \text{и}\ h(t)=\frac{f(t)}{1-F(t)},$$

поэтому если в момент a_j число отказов равно d_j, то

$$h_j=h(a_j)=\frac{d_j}{N_j},\quad\text{где}\quad N_j=\sum_{i=j}^{n}d_i.$$

В случае, когда имеются приостановки l_j, то $h(a_j)=\dfrac{d_j}{N_j}$, где $N_j=\sum_{i=j}^{n}(d_i+l_i)$.

В непрерывном случае

$$h(t)=-\frac{d}{dt}\ln(1-F(t)).$$

Отсюда, обозначая $H(t)=\int_0^t h(x)\,dx$. Имеем:

1) $H(t)=-\ln(1-F(t))$ и 2) $F(t)=\exp(-H(t))$.

Оценкой для $H(t)$ служит $\sum_{t_j<t}h_j$, т.е. $H(t)\approx\sum_{t_j<t}h_j\approx-\sum_{t_j<t}\ln(1-h_j)$.

Значит,

$$F(t)=\exp(-H(t))\approx\exp\left(\sum_{t_j<t}\ln\,(1-h_j)\right)=\exp\left(\ln\prod_{t_j<t}(1-h_j)\right)=\prod_{t_j<t}(1-h_j)=\prod_{t_j<t}\left(1-\frac{d_j}{N_j}\right).$$

Теперь рассмотрим функцию правдоподобия. Она будет равна

$$L=\prod(h_j)^{d_j}(1-h_j)^{l_j}\text{ , т.е. }\quad \ln L=\sum(d_j\ln h_j+l_j\ln(1-h_j)).$$

Запишем оценку Каплана-Мейера в ином виде. Пусть $N_n^{(j)}=\#\left\{1\le k\le n: t_k<t_j\right\}$. Тогда оценка Каплана-Мейера будет определяться следующим соотношением:

$$\hat{S}_n(t)=1-\hat{F}_n(t)=\prod_{1\le j\le n:\,t_j<t}\left(\frac{n-N_n^{(j)}-1}{n-N_n^{(j)}}\right)^{\delta_j}=\prod_{j=1:\,t_j<t}\left(\frac{n-j}{n-j+1}\right)^{\delta_n^{(j)}},$$

где $\delta_n^{(j)}, 1\le j\le n,$ – порядковые статистики, построенные по δ_i, $i=1,2,\ldots,n$.

Рассмотрим случай дискретных величин и выведем отсюда формулу для оценки Каплана-Мейера функции времени жизни. Пусть D и L – дискретные случайные величины со значениями $d_1=l_1, d_2=l_2,\ldots,d_m=l_m$ и $T_i=\min(D_i,L_i)$, $\delta_i=I(L_i\ge D_i)$. Упорядочим наблюдения T по первой компоненте, т.е. рассмотрим вариационный ряд $T_{(1)}<T_{(2)}<\ldots<T_{(n)}$.

Определим события A_i следующим образом:

$$\delta_i=1\ \Leftrightarrow\ A_i=\{(d_i,l_j): l_j\ge d_i\};$$

$$\delta_i=0\ \Leftrightarrow\ A_i=\{(d_j,l_i): l_i<d_j\}.$$

Положим $p_i=P(A_i)$. Тогда функция правдоподобия будет равна

$$L(p_1,p_2,\ldots,p_n)=\ln P\left(\bigcap_{i=1}^{n}\bigcup_{j=i}^{m}[(D_i,L_j)\in A_i]\right)=\ln\prod_{i=1}^{n}p_i,$$

которое достигает максимума по p_i при $p_i=1/n$ для любого $i=1,2,\ldots,n$.

Рассмотрим пример:

$$0.8^+, 1.0, 2.7^+, 3.2, 5.4, 9.2^+, 12.1,$$

где + означает, что наблюдается величина L $(\delta=0)$, если + нет, то наблюдается величина D $(\delta=1)$.

Пусть $P(D_i=d_i, L_j=l_j)=p_i q_j$.

Если $\delta_i=1$, то мы имеем

$$\frac{1}{n}=P(A_i)=p_i(1-\sum\nolimits_{j<i}q_j);$$

если $\delta_i=0$, то мы имеем

$$\frac{1}{n}=P(A_i)=q_i(1-\sum\nolimits_{j\le i}p_j).$$

Для нашего примера имеем в таком случае систему уравнений:

$$\begin{cases} \dfrac{1}{7} = p_2(1-q_1), \\ \dfrac{1}{7} = p_4(1-q_1-q_3), \\ \dfrac{1}{7} = p_5(1-q_1-q_3), \\ \dfrac{1}{7} = p_7(1-q_1-q_3-q_6), \end{cases} \qquad \begin{cases} \dfrac{1}{7} = q_1, \\ \dfrac{1}{7} = q_3(1-p_2), \\ \dfrac{1}{7} = q_6(1-p_2-p_4-p_5). \end{cases}$$

Решая ее, получаем значения

$$p_2 = \frac{1}{6},\ p_4 = p_5 = \frac{5}{24},\ p_7 = \frac{5}{12},\ \ q_1 = \frac{1}{7},\ q_3 = \frac{6}{35},\ q_6 = \frac{12}{35}.$$

Покажем, что эти решения определяют оценку Каплана-Мейера. Действительно,

$$\hat{S}(1.0) = 1 - p_2 = 1 - \frac{1}{7-1} = \frac{5}{6},$$

$$\hat{S}(3.2) = (1 - p_2 - p_4) = \frac{15}{24} = \frac{5}{8} = \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{4}\right),$$

$$\hat{S}(5.4) = (1 - p_2 - p_4 - p_5) = \frac{10}{24} = \frac{5}{12} = \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{3}\right),$$

$$\hat{S}(12.1) = (1 - p_2 - p_4 - p_5 - p_7) = 0 = \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{4}\right)(1-1).$$

Покажем, что такие соотношения имеют место в общем случае. Пусть есть выборка:

$$1(n_1),\ 2^{+}(m_2),\ 3(n_3),\ 4^{+}(m_4)\ ,\quad n_1 + m_2 + n_3 + m_4 = n,$$

т.е. «1» встретилась n_1 раз, «2» – m_2 раз и т.д.

Здесь для определения вероятностей p_i и q_j имеем систему уравнений:

$$\begin{cases} \dfrac{n_1}{n} = p_1, & \dfrac{m_2}{n} = q_2(1-p_1), \\ \dfrac{n_3}{n} = p_3(1-q_2), & \dfrac{m_4}{n} = q_4(1-p_1-p_3). \end{cases}$$

Решение этой системы – это:

$$p_1 = \frac{n_1}{n},\ p_3 = \frac{n_3}{n}\cdot\frac{n-n_1}{n-n_1-m_2},\ q_2 = \frac{m_2}{n-n_1},\ q_4 = \frac{m_4(n-n_1-m_2)}{(n-n_1)(n-n_1-m_2-n_3)}.$$

Поэтому,

$$1 - p_1 = 1 - \frac{n_1}{n},$$

$$1 - (p_1 + p_3) = \left(1 - \frac{n_1}{n}\right)\left(1 - \frac{n_3}{n - n_1 - m_2}\right)$$ – оценка Каплана-Мейера для данной выборки.

Аналогично проводятся рассуждения в общем случае.

Чтобы определить **оценку Нельсона**, приведем один общий результат, касающийся аддитивных статистик.

Пусть $(X_j, Y_j), 1 \le j \le n,$ – последовательность независимых одинаково распределенных пар случайных величин, причем X_j и Y_j, $j = 1, 2, \ldots, n,$ независимы также между собой. Обозначим $F(x)$ и $G(x)$ функции распределения случайных величин X_1 и Y_1 соответственно. Введем величины $\varsigma_j = \min(X_j, Y_j)$ и $w_j = I(X_j < Y_j)$.

Если $H(x) = \mathbf{P}(\varsigma_j < x)$, то $1 - H(x) = (1 - F(x))(1 - G(x))$.

Рассмотрим статистику

$$W_n = \frac{1}{n} \sum_{j=1}^{n-1} \frac{a(\varsigma_n^{(j)}) \cdot w_n^{[j]}}{1 - j/n},$$

где $a(x)$ – измеримая функция, $\varsigma_n^{(j)}$ – j-я порядковая статистика, а $w_n^{[j]}$ – индуцированная порядковая статистика: пусть пары $(\varsigma_i, \delta_i), 1 \le i \le n,$ упорядочены по первой компоненте и если $\varsigma_n^{(j)} = \varsigma_i$, то $\delta_n^{[j]} = \delta_i$.

Оценка Нельсона определяется следующим образом:

$$V_n(x) = \frac{1}{n} \sum_{j=1}^{n-1} \frac{I(\varsigma_n^{(j)} < x) \cdot \delta_n^{[j]}}{1 - j/n}.$$

Можно показать, что $V_n(x) \underset{n\to\infty}{\overset{p}{\to}} \ln(1 - F(x)) = \ln S(x)$, а $\hat{S}_n(t) \underset{n\to\infty}{\overset{p}{\to}} S(t)$.

Более того, последовательность $\sqrt{n}\,(W_n - \mu)$ сходится по распределению к нормальной случайной величине с нулевым математическим ожиданием и дисперсией $\sigma^2 = \int\limits_{-\infty}^{+\infty} \frac{a^2(x)\,dF(x)}{(1 - F(x))^2 (1 - G(x))}$. Здесь $\mu = \int\limits_{-\infty}^{+\infty} \frac{a(x)\,dF(x)}{1 - F(x)}$.

Статистика $$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{a(\varsigma_n^{(i)})\,\delta_n^{[i]}}{\left(1 - \frac{i}{n}\right)^2}$$

асимптотически нормальна с математическим ожиданием, равным

$$\mu_S = \int\limits_{-\infty}^{+\infty} \frac{a^2(x)\,dF(x)}{(1 - F(x))^2 (1 - G(x-))},$$

поэтому ее можно использовать для оценки дисперсии оценки Нельсона.

Поскольку предельная дисперсия предлагаемых оценок зависит от неизвестной функции распределения, то для построения доверительных интервалов на функцию распределения стараются подобрать такое преобразование оценок, чтобы его предельная дисперсия не зависела от неизвестной функции распределения. Эти преобразования называются преобразованиями, стабилизирующими дисперсию.

Преобразование, стабилизирующее дисперсию.

Пусть распределение величин Z_n, $n=1, 2, \ldots$, зависит от параметра θ, $\{a_n\}$ – неограниченная числовая последовательность, не зависящая от θ. Пусть при $n\to\infty$ последовательность $a_n(Z_n-\theta)$ сходится по распределению к Z. Тогда

$$a_n(g(Z_n)-g(\theta)) \underset{n\to\infty}{\overset{d}{\to}} g'(\theta)Z$$

для любой непрерывно дифференцируемой функции $g(\theta)$ такой, что $g'(\theta)\neq 0$ Пусть $\mathbf{D}(Z)=\sigma^2(\theta)$, тогда $\mathbf{D}(g'(\theta)Z)=(g'(\theta)\sigma(\theta))^2$. Если подобрать такую функцию $g(x)$, что

$$g'(\theta)\sigma(\theta)=c, \tag{9}$$

c – постоянная, то асимптотическая дисперсия $g(Z_n)$ не будет зависеть от θ. Преобразования $g(Z_n)$, для которых выполняется соотношение (9), называются ***преобразованиями, стабилизирующими дисперсию***.

Задача. Пусть $X_1, X_2, \ldots, X_n$ независимы и имеют общее биномиальное распределение $b(1,\theta)$ (распределение Бернулли), $\bar{X}=\sum_{i=1}^n X_i/n$.

Показать, что: 1) $g(\bar{X})=\arcsin\sqrt{\bar{X}}$ является преобразованием, стабилизирующим дисперсию: 2) $L\ \{\sqrt{n}\left(\arcsin\sqrt{\bar{X}}-\arcsin\sqrt{\theta}\right)\Rightarrow N(0,1/4)$.

Заметим также, что $\left(\arcsin\sqrt{x}\right)'=\dfrac{1}{2\sqrt{x(1-x)}}$.

$$\mathbf{D}(\arcsin\sqrt{\hat{S}(t)})\approx\frac{\mathbf{D}(\hat{S}(t))}{4\hat{S}(t)(1-\hat{S}(t))}=\frac{\hat{S}^2(t)\sigma_S^2(t)}{4\hat{S}(t)(1-\hat{S}(t))}\ \Rightarrow\ \sqrt{\mathbf{D}(\arcsin\sqrt{\hat{S}(t)})}=\frac{\sigma_S(t)}{2}\sqrt{\frac{\hat{S}(t)}{1-\hat{S}(t)}}.$$

Далее,

$$\mathbf{D}(\ln\hat{S}(t))\approx\frac{\mathbf{D}(\hat{S}(t))}{\hat{S}^2(t)}=\sigma_S^2(t)\ \Rightarrow\ \frac{\ln\hat{S}(t)-\ln S(t)}{\sqrt{\mathbf{D}(\ln\hat{S}(t))}}\approx\frac{\ln\hat{S}(t)-\ln S(t)}{\sigma_S(t)}.$$

Доверительный интервал имеет вид $-C\le\dfrac{\ln\hat{S}(t)-\ln S(t)}{\sigma_S(t)}\le C$. Поэтому нижняя доверительная граница равна

$$\ln L=\ln\hat{S}(t)-C\sigma_S(t)=\ln\hat{S}(t)\left(1-\frac{C\sigma_s(t)}{\ln\hat{S}(t)}\right)\le\ln\hat{S}(t)\exp\left(-\frac{C\sigma_s(t)}{\ln\hat{S}(t)}\right)=\frac{1}{\theta}\ln\hat{S}(t)=\ln((\hat{S}(t))^{1/\theta}).$$

Отсюда $L=(\hat{S}(t))^{1/\theta}$.

Рассмотрим следующее тестовое задание.

Тест 1. Пусть $F(x)=\mathbf{P}(X \le x)$, $S(x)=1-F(x)$. Имеется следующая случайно цензурированная выборка объема 20:

$$1,\ 2,\ 3^+,\ 4,\ 4,\ 4^+,\ 4^+,\ 5,\ 7^+,\ 8,\ 8,\ 8,\ 9,\ 9,\ 9,\ 9,\ 10^+,\ 10^+,\ 12,\ 12.$$

Здесь наблюдения, помеченные знаком $+$ означают приостановку (т.е. в модели $(w_j,\zeta_j),\ \zeta_j=\min(X_j,Y_j)$ наблюдается величина Y_j), а без плюса – приостановку (т.е. наблюдается величина X_j).

Задания.

1. Найдите $\hat{S}(11)-\tilde{S}(11)$ (Kaplan-Meier'а оценка – Nelson-Aalen'а оценка).

A) -0.083 **B)** -0.076 **C)** 0 **D)** 0.074 **E)** 0.035.

2. Пусть 12 – последнее наблюдаемое значение. Оцените $S(20)$ по методу Brown, Hollander, Kowar.

A) 0.09 **B)** 0.11 **C)** 0.22 **D)** 0.27 **E)** 0.32.

3. Найдите $\hat{V}(\hat{S}(11))-\hat{V}(\tilde{S}(11))$ (Greenwood – Aalen-Johansen'а оценки)

A) -0.025 **B)** -0.005 **C)** -0.002 **D)** 0.005 **E)** 0.015.

4. Пусть $C-$нижняя граница 95%-го доверительного интервала для $S(11)$; $D-$ нижняя граница 95%-го доверительного интервала для $\ln S(11)$. Найдите $C-D$ (с точностью до 0.005).

A) 0.01 **B)** 0 **C)** -0.04 **D)** -0.05 **E)** -0.06.

5. Определите нижнюю границу 95%-го доверительного интервала для величины $\arcsin\sqrt{S(11)}$.

A) 0.01 **B)** 0.04 **C)** 0.09 **D)** 0.12 **E)** 0.12.

6. Найдите оценку $\hat{\mu}$ среднего времени жизни.

A) 7.6 **B)** 7.8 **C)** 8.0 **D)** 8.1 **E)** 8.4.

7. Найдите оценку $\hat{V}(\hat{\mu})$ дисперсии среднего времени жизни

A) 0.54 **B)** 0.57 **C)** 0.59 **D)** 0.61 **E)** 0.65.

8. Найдите число моментов времени из множества $t=\{4,5,8,9\}$, попавших в 99%-й доверительный интервал для квантиля порядка 0.75 распределения $F(x)$.

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4.

Решение.

1. Классическая оценка Каплана-Мейера имеет вид:

$$\hat{S}(t)=\prod_{\varsigma_i \le t}\left(1-\frac{1}{Y_i}\right)^{\delta_i}, \text{ где } \quad Y_1 = N,\ Y_{i+1}=Y_i-d_i-c_i.$$

Для повторяющихся наблюдений можно применять следующую формулу:

$$\hat{S}(t)=\prod_{\varsigma_i \le t}\left(1-\frac{d_i}{Y_i}\right)^{\delta_i}.$$

Для данной выборки имеем:

$Y_1=20 \qquad Y_2=Y_1-d_1-c_1=20-1-0=19,$

$Y_3=17 \qquad Y_4=13 \qquad Y_5=11 \qquad Y_6=8 \qquad Y_7=2$

$d_1=1 \qquad d_2=1 \qquad d_3=2 \qquad d_4=1 \qquad d_5=3 \qquad d_6=4 \qquad d_7=2$

$c_1=0 \qquad c_2=1 \qquad c_3=2 \qquad c_4=1 \qquad c_5=0 \qquad c_6=2 \qquad c_7=0$

Формально, мы считаем, что исходная случайная величина имеет непрерывное распределение и несколько «раздвигая» повторяющиеся наблюдения, можно обосновать данную формулу.

Для представленного примера получаем:

$$\hat{S}(1)=\left(1-\frac{1}{20}\right)=0.95,$$

$$\hat{S}(2)=\left(1-\frac{1}{20}\right)\left(1-\frac{1}{19}\right)=(0.95)\left(1-\frac{1}{19}\right)=0.90,$$

$$\hat{S}(4)=\left(1-\frac{1}{20}\right)\left(1-\frac{1}{19}\right)\left(1-\frac{2}{17}\right)=(0.90)\left(1-\frac{2}{17}\right)=0.7941176470\approx 0.79412,$$

$$\hat{S}(5)=(0.79412)\left(1-\frac{1}{13}\right)\approx 0.73303,$$

$$\hat{S}(8)=(0.73303)\left(1-\frac{3}{11}\right)=0.5331139449\approx 0.53311,$$

$$\hat{S}(9)=(0.533311)\left(1-\frac{4}{8}\right)=0.2665569\approx 0.26656,$$

$$\hat{S}(12)=(0.26656)\left(1-\frac{2}{2}\right)=0.$$

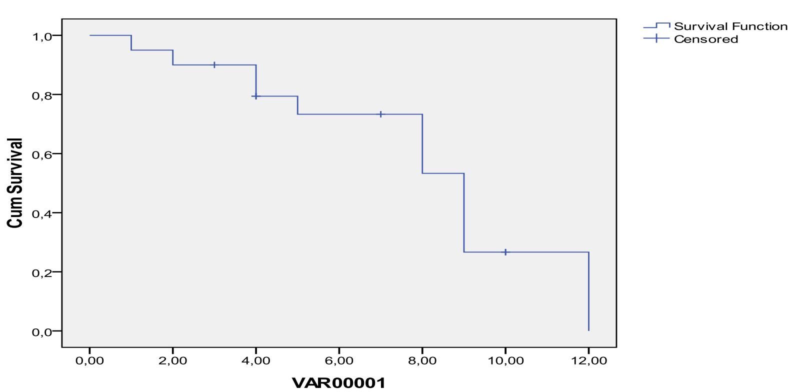
Для всех $9\le t<12$ имеем: $\hat{S}(t)=0.26656$, поэтому $\hat{S}(11)=0.26656$.

Ниже приведены вычисления в пакете SPSS:

Survival Table

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	1,000	1,00	,950	,049	1	19
2	2,000	1,00	,900	,067	2	18
3	3,000	,00	.	.	2	17
4	4,000	1,00	.	.	3	16
5	4,000	1,00	,794	,092	4	15
6	4,000	,00	.	.	4	14
7	4,000	,00	.	.	4	13
8	5,000	1,00	,733	,103	5	12
9	7,000	,00	.	.	5	11
10	8,000	1,00	.	.	6	10
11	8,000	1,00	.	.	7	9
12	8,000	1,00	,533	,124	8	8
13	9,000	1,00	.	.	9	7
14	9,000	1,00	.	.	10	6
15	9,000	1,00	.	.	11	5
16	9,000	1,00	,267	,113	12	4
17	10,000	,00	.	.	12	3
18	10,000	,00	.	.	12	2
19	12,000	1,00	.	.	13	1
20	12,000	1,00	,000	,000	14	0

Survival Function



Далее, $$\tilde{H}(t)=\sum_{\varsigma_i\le t}\frac{\delta_i}{Y_i},$$
и, как и ранее, будем использовать следующую формулу:
$$\tilde{H}(t)=\sum_{\varsigma_i\le t}\frac{d_i}{Y_i}.$$
Для нашего примера
$$\tilde{H}(11)=\frac{1}{20}+\frac{1}{19}+\frac{2}{17}+\frac{1}{13}+\frac{3}{11}+\frac{4}{8}=1.069928987,$$
$$\tilde{S}(11)=\exp(-\tilde{H}(11))=0.3430328763\approx 0.34303.$$
Таким образом,
$$\hat{S}(11)-\tilde{S}(11)=-0.07647.$$ **<u>Ответ:</u> B.**

2. $t_{\max}=12$. Для $t>t_{\max}$ будем использовать следующую формулу: $\hat{S}(t)=(\hat{S}(t_{\max}-0))^{t/t_{\max}}$, отсюда,
$$\hat{S}(20)=(\hat{S}(12-0))^{20/12}=(0.26656)^{5/3}=0.110405734.$$ **<u>Ответ:</u> B.**

3. *Greenwood:* $\hat{V}(\hat{S}(t))=\hat{S}^2(t)\cdot\sum_{\varsigma_i\le t}\frac{d_i}{Y_i(Y_i-d_i)}$:

$$\hat{V}(\hat{S}(1))=\frac{1}{20\cdot 19}\cdot(0.95)^2=0.00263\cdot(0.95)^2=0.00238,$$
$$\hat{V}(\hat{S}(2))=\left(0.00263+\frac{1}{19\cdot 18}\right)\cdot(0.9)^2=0.00556\cdot(0.9)^2=0.0045,$$
$$\hat{V}(\hat{S}(4))=\left(0.00556+\frac{2}{17\cdot 15}\right)\cdot(0.79412)^2=0.01340\cdot(0.79412)^2=0.00845,$$
$$\hat{V}(\hat{S}(5))=\left(0.01340+\frac{1}{13\cdot 12}\right)\cdot(0.73303)^2=0.01981\cdot(0.73303)^2=0.01064,$$
$$\hat{V}(\hat{S}(8))=\left(0.01981+\frac{3}{11\cdot 8}\right)\cdot(0.53311)^2=0.05390\cdot(0.53311)^2=0.01532,$$
$$\hat{V}(\hat{S}(9))=\left(0.05390+\frac{4}{8\cdot 4}\right)\cdot(0.26656)^2=0.17890\cdot(0.26656)^2=0.01271,$$
$$\hat{V}(\hat{S}(11))=0.01271.$$

Aalen-Johansen: $\tilde{V}(\hat{S}(t))=(\hat{S}(t))^2\sum_{\varsigma_i\le t}\frac{d_i}{Y_i^2}$:

$$\tilde{V}(\hat{S}(1))=\frac{1}{20^2}\cdot(0.95)^2=(0.0025)\cdot(0.95)^2=0.00225625\approx 0.00226,$$
$$\tilde{V}(\hat{S}(2))=\left(0.0025+\frac{1}{19^2}\right)\cdot(0.9)^2=(0.00527)\cdot(0.9)^2=0.00427,$$
$$\tilde{V}(\hat{S}(4))=\left(0.00527+\frac{2}{17^2}\right)\cdot(0.79412)^2=(0.01219)\cdot(0.79412)^2=0.00769,$$

$$\tilde{V}(\hat{S}(5)) = \left(0.01219 + \frac{1}{13^2}\right) \cdot (0.73303)^2 = (0.01811) \cdot (0.73303)^2 = 0.00973,$$

$$\tilde{V}(\hat{S}(8)) = \left(0.01811 + \frac{3}{11^2}\right) \cdot (0.53311)^2 = (0.04290) \cdot (0.53311)^2 = 0.01219,$$

$$\tilde{V}(\hat{S}(9)) = \left(0.04290 + \frac{4}{8^2}\right) \cdot (0.26656)^2 = (0.1054) \cdot (0.26656)^2 = 0.00749,$$

$$\tilde{V}(\hat{S}(11)) = 0.00749,$$

$$\hat{V}(\hat{S}(11)) - \tilde{V}(\hat{S}(11)) = 0.01271 - 0.00749 = 0.00522.$$

Ответ: D.

4. Имеем: $\sigma_S^2(11) = \sum_{\varsigma_i \le 11} \frac{d_i}{Y_i(Y_i - d_i)} = 0.1789$ (из задания 3),

$$\sigma_S(11) = 0.42297,$$

$$\hat{S}(11) = 0.26656,$$

$$\theta = \exp\left(\frac{(1.96)\cdot\sigma_S(11)}{\ln \hat{S}(11)}\right) = 0.5312,$$

$$C = \hat{S}(11) - (1.96)\cdot\sigma_S(11)\cdot\hat{S}(11) = 0.045578,$$

$$D = \left(\hat{S}(11)\right)^{1/\theta} = (0.26656)^{1/0.5342} = 0.08415,$$

$$C - D = 0.045578 - 0.08415 = -0.03858 \approx -0.04.$$

Ответ: C.

5. Имеем: $\arcsin((\hat{S}(11))^{1/2}) - (0.5)\cdot z_{0.975}\cdot\sigma_S(11)\cdot\left(\frac{\hat{S}(11)}{1-\hat{S}(11)}\right)^{1/2} =$

$$= 0.26656 - (0.5)\cdot(1.96)\cdot(0.42297)\cdot\left(\frac{0.26656}{1-0.26656}\right)^{1/2} = 0.2926.$$

Нижняя доверительная граница интервала равна
$\sin^2\left(\max(0, 0.2926)\right) = 0.0832.$

Ответ: B.

6. Оценка $\hat{S}(t)$ Каплана-Мейера имеет вид:

t	$[0,1)$	$[1,2)$	$[2,4)$	$[4,5)$	$[5,8)$	$[8,9)$	$[9,12)$
$\hat{S}(t)$	1	0.95	0.9	0.79412	0.73303	0.53311	0.26656

поэтому оценка среднего времени равна

$$\hat{\mu} = 1\cdot(1-0) + 0.95\cdot(2-1) + 0.9\cdot(4-2) + 0.79412\cdot(5-4) + 0.73303\ (8-5) + 0.53311\cdot(9-8) +$$
$$+ 0.26656\cdot(12-9) = 8.076.$$

Ответ: D.

Эту же оценку можно вычислить так (по исходному определению математического ожидания):

$$\hat{\mu} = 1\cdot 0.05 + 2\cdot 0.05 + 4\cdot 0.10588 + 5\cdot 0.06109 + 8\cdot 0.19992 + 9\cdot 0.26655 + 12\cdot 0.26656 = 8.076.$$

7. Составим следующую таблицу:

t_i	1	2	4	5	8	9
$\int_{t_i}^{12} \hat{S}(t)\,dt$	7.076	6.126	4.326	3.53188	1.33278	0.79967
$\dfrac{d_i}{Y_i(Y_i-d_i)}$	0.00263	0.00292	0.00784	0.00641	0.03409	0.125

Тогда

$$\hat{V}(\hat{\mu}) = \sum_{i=1}^{6}\left(\int_{t_i}^{12}\hat{S}(t)\,dt\right)^2 \cdot \frac{d_i}{Y_i(Y_i-d_i)} = (7.076)^2\cdot(0.00263) + (6.126)^2\cdot(0.00292) +$$

$$+(4.326)^2\cdot(0.00784) + (3.53188)^2\cdot(0.00641) + (1.33279)^2\cdot(0.03409) + (0.79968)^2\cdot(0.125) = 0.6087.$$

<u>Ответ:</u> D.

Приведем вычисления в пакете SPSS:

Means and Medians for Survival Time

Mean[a]				Median			
		95% Confidence Interval				95% Confidence Interval	
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
8,076	,810	6,489	9,663	9,000	,967	7,105	10,895

a. Estimation is limited to the largest survival time if it is censored.

8. Оценка квантиля порядка 0.75 равна $\hat{x}_{0.75} = \min\{t : \hat{S}(t) \le 1-0.75\} = 12$. Точка t принадлежит 99%-му доверительному интервалу для квантиля $x_{0.75}$ порядка 0.75, если для нее выполняется неравенство:

$$-2.576 \le \frac{\hat{S}(t)-0.25}{\sqrt{\hat{V}(\hat{S}(t))}} \le 2.576.$$

Таким образом,

t	4	5	8	9
$\hat{S}(t)$	0.79412	0.73303	0.53311	0.26656
$\hat{V}(\hat{S}(t))$	0.00845	0.01064	0.01532	0.01271
$\dfrac{\hat{S}(t)-0.25}{\sqrt{\hat{V}(\hat{S}(t))}}$	5.9192	4.6819	2.2873	0.1469

<u>Ответ:</u> C.

Пусть $\hat{S}(t)$ – оценка для функции $S(t)$. Рассмотрим функцию $\ln \hat{S}(t)$ и произведем ее разложение Тейлора. Имеем:

$$\ln \hat{S}(t) = \ln S(t) + \frac{(\hat{S}(t) - S(t))}{S(t)} + o(\hat{S}(t) - S(t)).$$

Тогда для дисперсии оценки $\ln \hat{S}(t)$ получаем:

$$\mathbf{D}(\ln \hat{S}(t)) \sim \frac{\mathbf{D}(\hat{S}(t))}{S^2(t)}.$$

Заменим функцию $S(t)$ ее оценкой, поэтому перепишем предыдущее равенство следующим образом:

$$\mathbf{D}(\ln \hat{S}(t)) \sim \frac{\mathbf{D}(\hat{S}(t))}{\hat{S}^2(t)}.$$

В качестве оценки для дисперсии $\mathbf{D}(\hat{S}(t))$ довольно часто используют следующую оценку:

$$\sum_i \frac{d_i}{Y_i(Y_i - d_i)} \quad \text{(Greenwood)}$$

Отметим, что КМ-оценки были получены в 1958 г., а формула Greenwood в 1924г.

Выведем формулу Greenwood. В случае полной выборки пусть $\hat{p} = \frac{x}{n}$ – частота. Тогда

$$\mathbf{D}(\hat{p}) = \frac{p(1-p)}{n} \approx \frac{1}{n} \cdot \frac{x}{n}\left(1 - \frac{x}{n}\right) = \frac{x(n-x)}{n^3}.$$

Если t таково, что $S(t) = p$, то

$$\hat{S}(t) = \hat{p} = \frac{n(t)}{n(t) + d(t)} = \frac{n(t)}{n(0)}.$$

Рассмотрим теперь Greenwood-оценку в случае цензурирования. Имеем:

$$\mathbf{D}(\hat{S}(t)) \approx \hat{S}^2(t)\sum_{t_i \le t}\frac{d_i}{n_i(n_i - d_i)} = \hat{S}^2(t)\sum_{t_i \le t}\frac{n_{i+1} - n_i}{n_{i+1}n_i} = \hat{S}^2(t)\sum_{t_i \le t}\left(\frac{1}{n_{i+1}} - \frac{1}{n_i}\right) = \hat{S}^2(t)\left(\frac{1}{n_j} - \frac{1}{n_0}\right) =$$

$$= \hat{S}^2(t)\frac{d(t)}{n(t)n(0)} = \frac{\hat{S}^2(t)(1 - \hat{S}(t))}{n(t)} = \frac{1}{n(t)}\frac{d(t)}{n(0)}\frac{n(t)}{n(0)}\frac{n(t)}{n(0)} = \frac{d(t)n(t)}{n^3(0)}.$$

С другой стороны,

$$\mathbf{D}(\ln \hat{S}(t)) = \sum_{t_i \le t}\mathbf{D}\left(\ln\left(1 - \frac{d_i}{Y_i}\right)\right) \approx \sum_{t_i \le t}\left(1 - \frac{d_i}{Y_i}\right)^{-2}\mathbf{D}\left(1 - \frac{d_i}{Y_i}\right) = \sum_{t_i \le t}\left(1 - \frac{d_i}{Y_i}\right)^{-2}\mathbf{D}\left(\frac{d_i}{Y_i}\right) =$$

$$\sum_{t_i \le t}\left(1 - \frac{d_i}{Y_i}\right)^{-2}\frac{1}{Y_i}\frac{d_i}{Y_i}\left(1 - \frac{d_i}{Y_i}\right) = \sum_{t_i \le t}\frac{d_i}{Y_i(Y_i - d_i)}.$$

Рассмотрим теперь подробнее оценки для математического ожидания величины X. Для оценки математического ожидания можно пользоваться одной из двух формул – прямое определение, данное как интеграл $\int\limits_0^\infty x\,dF(x)$, и как следующий интеграл $\int\limits_0^\infty (1-F(x))\,dx$ (см. Гнеденко Б.В. Курс теории вероятностей, М: Едиториал УРСС, 2005 с.151):

$\hat{\mu} = \int\limits_0^\infty (1-\hat{F}_n(x))\,dx = \int\limits_0^\infty x\,d\hat{F}_n(x)$, поскольку если величина неотрицательна, то ее математическое ожидание равно

$$\mu = \int\limits_0^\infty x\,dF(x) = -\int\limits_0^\infty x\,dS(x) = -\left. xS(x)\right|_0^\infty + \int\limits_0^\infty S(x)\,dx = \int\limits_0^\infty S(x)\,dx = \int\limits_0^\infty (1-F(x))\,dx.$$

Далее, для полной выборки

$$\int\limits_0^\infty [\int\limits_x^\infty S(u)\,du]^2 \frac{dF(x)}{S^2(x)} = -\int\limits_0^\infty [\int\limits_x^\infty S(u)\,du]^2 \frac{dS(x)}{S^2(x)} = \int\limits_0^\infty [\int\limits_x^\infty S(u)\,du]^2 d\left(\frac{1}{S(x)}\right) = \left.[\int\limits_x^\infty S(u)\,du]^2 \cdot \frac{1}{S(x)}\right|_0^\infty +$$

$$+\int\limits_0^\infty \frac{2}{S(x)} \int\limits_x^\infty S(u))\,du \cdot S(x)\,dx = -\mu^2 - \left. 2x\int\limits_x^\infty S(u)\,du\right|_0^\infty + 2\int\limits_0^\infty xS(x)\,dx = -\mu^2 + \int\limits_0^\infty S(x)\,d\,(x^2) =$$

$$= -\mu^2 + \int\limits_0^\infty x^2\,dF(x).$$

Поэтому для оценки дисперсии мы будем использовать первый интеграл с учетом специфики цензурированных выборок.

Далее мы приведем практический пример расчета среднего времени нахождения финансовых активов в банках. Пример взят из книги Рейли Р., Швайс Р. Оценка нематериальных активов. М.: ИД «Квинто-консалтинг», 2005 – 772 с.

4. Анализ срока службы и выбытия финансовых активов.

Анализ срока службы — это изучение условий размещения активов (или инвестиций в них) и их последующего выбытия для установления характеристик срока службы этих активов. В частности, важными характеристиками срока службы являются средний срок службы и схема выбытия активов, которую иногда называют темпом изнашивания. На основе оценки этих характеристик (срока службы и информированного суждения о текущей и вероятной будущей операционной среде) аналитик может оценить остаточный срок полезного использования действующих активов. Рассмотрим какую-либо финансовую контору (это может быть банк, брокерская фирма и т.д.).

В табл. 1 представлен иллюстративный набор данных анализа срока службы нематериального актива. В ней представлены данные по возрасту и сроку службы, относящиеся к списку клиентов брокерской фирмы, занимающейся операциями с ценными бумагами. В этом примере мы имеем историю всех отношений с оставшимися клиентами для каждой группы «одного возраста» — или группы размещения клиентских счетов — за период с 1995 г. по 2006 г. Например, из 29 клиентских счетов, открытых в 1997 г. — т.е., из группы клиентов, появившихся в 1997 г. или группы счетов, размещенных в этом году — 21 клиент продолжил отношения до 1998 г. Из этой же самой группы 12 клиентов продолжили отношения до 1999 г., 6 клиентов сохранили отношения до 2000 г. и 4 клиента продолжили отношения до 2001 г. В нижней строке таблицы 1 суммируется общее число оставшихся клиентов из каждой группы «одного возраста», а также общее число выбывших клиентов (т.е., тех клиентов, которые прекратили вести бизнес с данным брокером) из каждой группы. Эта нижняя строка данных по возрасту и сроку службы может быть единственной информацией о сроках службы нематериальных активов, имеющейся в наличии для исследования во многих случаях проведения анализа нематериальных активов.

Для построения кривой выживаемости, используемой в аналитическом методе определения срока службы, мы начнем с опытного интервала определенной ширины, скажем, с трехлетнего периода с 2003 по 2005 гг. (т.е., годом оценки является 2006 г.) Ширина опытного интервала может впоследствии измениться.

Далее мы выполним анализ чувствительности для оценки «пригодности для аппроксимации» различных аналитических кривых, которые будут сравниваться с фактической кривой выживаемости оцениваемого нематериального актива. Следует отметить, что опытный интервал соответствует конкретным годам до даты оценки, за которые будут подвергнуты всестороннему анализу данные о сроке службы. Мы выбрали опытный интервал, включающий последние три года до даты оценки — т.е., 2003 г., 2004 г. и 2005 г. Этот опытный интервал представлен затененными столбцами, соответствующими 2003 г., 2004 г. и 2005 г.

ТАБЛИЦА 1. Аналитический метод анализа срока службы. Пример построения кривой выживаемости. Данные о возрасте и сроке службы клиентских счетов. Размещение и выбытие клиентских счетов.
1. В каждой клетке верхнее число показывает «ПОТЕНЦИАЛЬНОЕ ВЫБЫТИЕ», а нижнее число показывает «ФАКТИЧЕСКОЕ ВЫБЫТИЕ» в течение года.
2. Горизонтальные строки соответствуют интервалу размещений.
3. Вертикальные столбцы соответствуют опытному интервалу.

В качестве точки сравнения, интервалы размещений соответствуют анализу размещения счетов за прошлые периоды. В сравнении с опытными интервалами (представленными в *столбцах* данных о возрасте и сроке службы), интервалы размещений представлены данными в *строках,* относящихся к исследуемым группам «одного возраста», представленным в табл. 1.

Таблица 1

Начало размещения	**Опыт**											(остаток на 2006 г.
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1995 95	95 0	95 9	86 19	67 6	61 16	45 4	41 3	38 1	37 9	28 0	28 3	25
1996 21		21 6	15 5	10 2	8 1	7 1	6 1	5 0	5 1	4 0	4 1	3
1997 29			29 8	21 9	12 6	6 2	4 0	4 0	4 2	2 0	2 0	2
1998 23				23 3	20 10	10 1	9 2	7 1	6 1	5 0	5 2	3
1999 14					14 7	7 1	6 1	5 1	4 2	2 0	2 0	2
2000 41						41 10	31 2	29 3	26 12	14 0	14 2	12
2001 63							63 17	46 9	37 15	22 3	19 3	16
2002 39								39 15	24 17	7 0	7 1	6
2003 23									23 17	6 0	6 0	6
2004 27										27 2	25 1	24
2005 35											35 6	29
2006 60												60
Итого 470 Остаток Выбытие	95 0	116 15	130 32	121 20	115 40	116 19	160 26	173 30	166 76	117 5	147 19	188

После того, как мы выбрали надлежащую ширину опытного интервала, подлежащего анализу, мы производим сортировку и группировку данных об активных и выбывших клиентских счетах по аналогичным возрастным группам (также называемых возрастными интервалами).

Из данных табл. 1 мы видим, что 4 клиента из группы клиентов, разместивших счета в 1999 г., 14 клиентов из группы клиентов, разместивших счета в 2000 г., и 19 клиентов из группы клиентов, разместивших счета в 2001 г., имеют четырехлетний «возраст» в 2003 г., 2004 г. и в 2005 г. соответственно. Это является релевантным фактом, поскольку период с 2003 по 2005 гг. представляет собой опытный интервал, выбранный нами для подробного анализа.

Аналогичным образом мы можем увидеть, что из групп 1999–2001 гг. в возрастном интервале от 4 до 5 лет выбыло соответственно 2, 0 и 3 клиента. Опять же, возрастным интервалом от 4 до 5 лет для клиентов, разместивших счета в 1999 г., является 2003 г., 2004 г. для клиентов, разместивших счета в 2000 г., и 2005 г.. для клиентов, разместивших счета в 2001 г. Сложив количество клиентских счетов, размещенных в 1999 г., 2000 г. и 2001 г., мы получим 37 (4 + 14 + 19 = 37) клиентских счетов в «возрасте» четырех лет в опытном интервале с 2003 г. по 2005 г. Из 37 счетов, достигших четырехлетнего «возраста», 5 (2 + 0 + 3 = 5) клиентских счетов выбыли (т.е. были закрыты) в течение четвертого года службы. Общее количество (37) клиентских счетов, которые по-прежнему были открыты в начале четвертого года своей службы — или количество клиентских счетов, которые потенциально могли выбыть в течение четвертого года службы — показано в столбце [2] и в строке для возрастного интервала 4–5.

Аналогичным образом, количество фактически выбывших (т.е. закрытых) клиентских счетов в возрастном интервале 4–5 (т.е.,5 счетов в «возрасте» 4 лет, которые были закрыты в течение четвертого года службы) приведено в столбце [3] табл. 2.

ТАБЛИЦА 2. Пример построения кривой выживаемости. Расчет выбранного опытного интервала (2003–2005 г.) кривой выживаемости с использованием метода темпа выбытия.

Таблица 2

Возрастной интервал	Кол-во счетов, которые потенциально могут выбыть	Кол-во фактически выбывших счетов в возрастном	Коэффициент выбытия	Коэффициент выживаемости	Кривая выживаемости
[1]	[2]	[3]	[4]	[5]	[6]
10–11	28	3	10,71%	89,29%	7,39%
9–10	32	1	3,13%	96,88%	8,28%
8–9	43	9	20,93%	79,07%	8,54%
7–8	12	3	25,00%	75,00%	10,80%
6–7	11	2	18,18%	81,82%	14,40%
5–6	22	3	13,64%	86,36%	17,61%
4–5	37	5	13,51%	86,49%	20,39%
3–4	55	16	29,09%	70,91%	23,57%
2–3	50	15	30,00%	70,00%	33,24%
1–2	55	18	32,73%	67,27%	47,49%
0 –1	85	25	29,41%	70,59%	70,59%
0	85				100,00%

Примечания:
Столбец [4] = столбец [3]/столбец [2]
Столбец [5] = 1 - столбец [4]
Столбец [6] = столбец [6] (из следующей строки) * столбец [5]

Путем сравнения количества счетов, выбывших в возрастном интервале, с количеством счетов, которые потенциально могли выбыть (иными словами, с количеством активных счетов на начало возрастного интервала), мы можем рассчитать коэффициент выбытия. Коэффициент выбытия иногда называют степенью выбытия. Доля единиц нематериальных активов — или, в данном примере, отношений с клиентами — остающихся активными, равна единице минус коэффициент выбытия. Таким образом, годовой коэффициент выбытия определяется следующим образом:

Коэффициент выбытия =Количество единиц, выбывших в течение периода/количество единиц, которые потенциально могли выбыть.

На основе годового коэффициента выбытия мы можем легко рассчитать годовой коэффициент выживаемости. Годовой коэффициент выживаемости определяется следующим образом:

Коэффициент выживаемости = 1 — Коэффициент выбытия.

Соответственно, для возрастного интервала клиентских счетов от 4 по 5 лет в выбранном опытном интервале с 2003 г. по 2005 г. мы имеем:

1. Коэффициент выбытия = 5 ÷ 37 = 0,1351, или 13,51%.
2. Коэффициент выживаемости = 1 — 0,1351 = 0,8649, или 86,49%.

В столбцах [4] и [5] табл. 2 представлены коэффициенты выбытия и выживаемости соответственно для различных возрастных интервалов клиентских счетов в выбранном опытном интервале с 2003 по 2005 гг.

Кривая выживаемости — это кумулятивная доля единиц нематериальных активов, остающихся активными по группе соответствующего предыдущего возрастного интервала. Например, для возрастного интервала от 4 до 5 лет, из последнего столбца табл. 2 мы видим, что в этом возрастном интервале потенциально могут выбыть 23,57 процентов клиентских счетов. Число 23,57 процентов соответствует проценту отношений с клиентами, действующих (или активных) в конце возрастного интервала от 3 до 4 лет. Из этих счетов «под риском» — 13,51 процентов выбывают, а 86,49 процентов «выживают» в следующем возрастном интервале (т.е., в интервале от четырех до пяти лет). Исходя из этого, мы имеем 20,39% процентов (т.е., 23,57 x 0,8649 = 20,39) отношений с клиентами, оставшихся в конце возрастного интервала от 4 до 5 лет.

Построенная кривая выживаемости начинается со 100 процентов действующих счетов в нулевом возрасте и доходит до *y* процентов действующих счетов в возрасте *x*. В табл. 2 построенная кривая выживаемости доходит до 7,39 процентов действующих счетов в возрастном интервале от 10 до 11 лет. По определению, это укороченная кривая выживаемости. В силу необходимости (поскольку еще не все счета закрыты) укороченная кривая выживаемости не доходит до точки нулевого процента действующих счетов.

Методы расчета кривых выживаемости.

Этот метод, использованный в примере для построения укороченной кривой выживаемости, называется методом годового коэффициента или темпа выбытия. Поскольку в этом методе используются как данные и информация о нематериальных активах, находящихся в эксплуатации, так и данные о выбывших нематериальных активах, он считается самым строгим из трех системных аналитических методов расчета кривых выживаемости. Другие два метода называются методом анализа исходной группы и методом анализа отдельной единицы.

В методе анализа исходной группы используются только данные о возрасте и сроке службы, относящиеся к использованию активов группы одного возраста. В методе анализа отдельной единицы, который отдельные аналитики считают методом, применимым в крайней случае, используются только данные о возрасте и сроке службы, относящиеся к выбывшим нематериальным активам (т.е., для этого должны использоваться данные о возрасте выбывших единиц нематериальных активов).

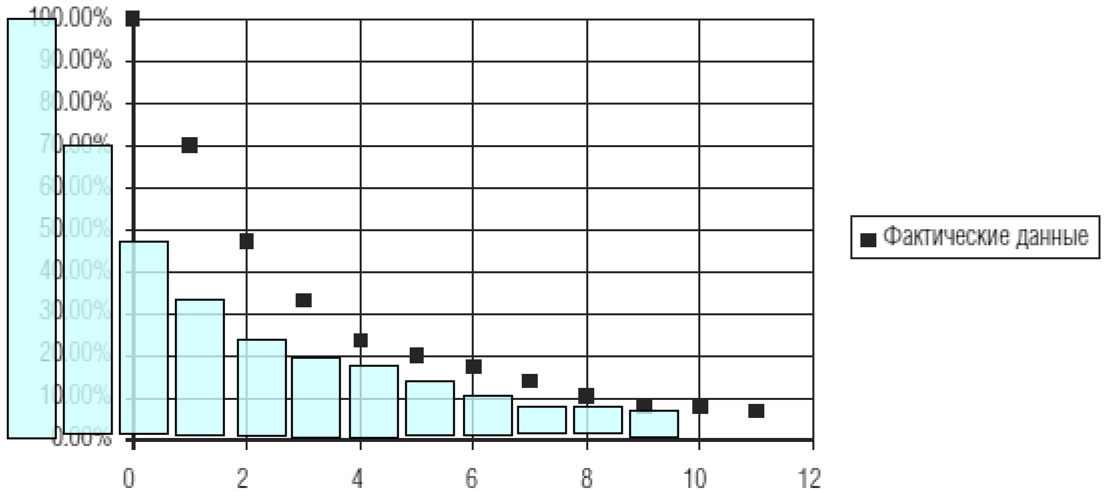


Рис. 3. Пример построения укороченной кривой выживаемости.

Установление среднего срока службы на основе кривых выживаемости.

На основе укороченной кривой выживаемости невозможно непосредственно оценить объективный средний срок службы или остаточный срок полезного использования, поскольку укороченная кривая не доходит до уровня нулевого процента действующих активов. В зависимости от количества проанализированных данных о возрасте и сроке службы, укороченная кривая может заканчиваться в любой точке между 99 процентами и 1 процентом действующих активов. Поэтому укороченная кривая выживаемости не позволяет произвести расчет общего срока службы исследуемой группы нематериальных активов. Тем не менее, отдельные наблюдения на основе укороченной кривой могут быть использованы для прогнозирования полной кривой путем сравнения построенной укороченной кривой с семейством стандартных кривых (функций) выживаемости. Именно, средний срок равен

$$1+0.71+0.47+0.33+0.24+0.2+0.18+0.14+0.11+0.09+0.08+0.07=3.62.$$

Выбор стандартной кривой (функции) выживаемости, наиболее подходящей для экстраполяции фактической укороченной кривой, часто основан на двух методах: «визуальный анализ» укороченной кривой или использование метода *наименьших квадратов,* т.е. статистического способа аппроксимации кривых. При визуальном анализе укороченной кривой эта кривая физически сравнивается с семейством кривых (функций) выживаемости стандартного типа для различных средних сроков службы.

В методе наименьших квадратов для аппроксимации кривых сначала вычисляют разность по вертикали между стандартной кривой выживаемости и фактической укороченной прямой в дискретных точках, а затем эта разница возводится в квадрат — то есть, умножается сама на себя. После этого суммируются квадраты разности между всеми точками наблюдения на стандартной кривой выживаемости и на фактической укороченной кривой. Стандартная кривая выживаемости, имеющая самый низкий или наименьший квадрат разности, считается кривой, наиболее репрезентативной в отношении характеристик срока службы рассматриваемого нематериального актива.

Поскольку стандартная кривая выживаемости достигает уровня нулевого процента действующих активов, можно вычислить площадь под кривой и, таким образом, достоверно оценить средний срок службы рассматриваемой группы нематериальных активов.

Имея достоверную оценку фактического среднего срока службы рассматриваемой группы нематериальных активов, на основе выбранной стандартной кривой выживаемости может быть вычислен средний остаточный срок службы.

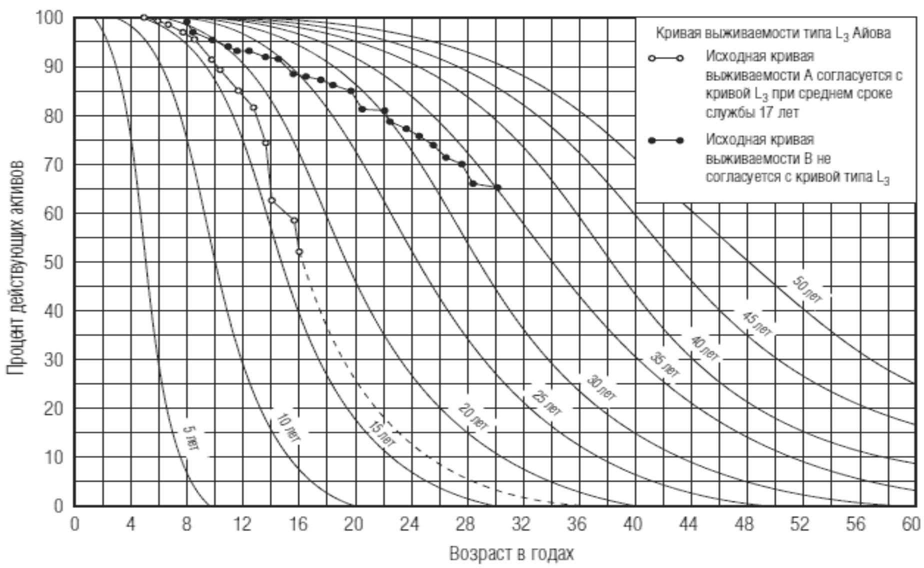


Рис. 4.1. Визуальный анализ укороченной кривой.

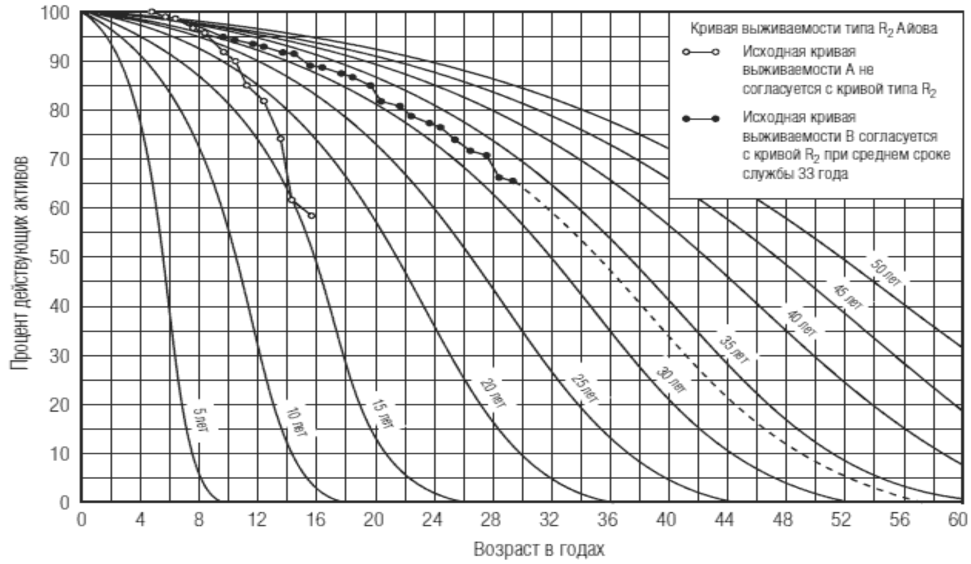


Рис. 4.2. Визуальный анализ укороченной кривой.

Источник: Anson Marston, Robley Winfrey, and Jean C. Hempstead, Engineering Valuation and Depreciation (Ames, IA: Iowa State University Press, 1953), стр. 166.

Стандартные кривые (функции) выживаемости.

Ниже представлены некоторые из наиболее распространенных типов стандартных кривых (функций) выживаемости, используемых в аналитическом методе анализа остаточного срока службы нематериальных активов:

1. Кривые типа Айова (Iova-type curves).
2. Распределение Вейбулла (Weibull) (сами кривые типа Айова могут быть частным случаем кривой выживаемости этого типа).
3. Кривые Гомпертца-Мейкхема (Gompertz-Makeham curve).
4. Полиномиальные функции.

Кривые типа Айова

Кривые выживаемости типа Айова были построены в результате исследований эмпирических данных, относящихся к характеристикам сохранивших работоспособность объектов промышленных и коммунальных предприятий многих различных видов. Эти данные включали эмпирические наблюдения в отношении: рельсовых шпал, крытых железнодорожных вагонов, телефонных столбов, станционного оборудования (например, телефонного оборудования), электрических трансформаторов и электропроводов, автомобилей, грузовых автомобилей коммерческого образца и так далее.

На рис. 5 изображена экспоненциальная кривая, аппроксимирующая построенную укороченную кривую. Экспоненциальная кривая трехлетнего среднего срока службы (exp. 3.0) обеспечивает лучшую аппроксимацию, чем exp. 3.5 на базе суммы квадратов разностей.

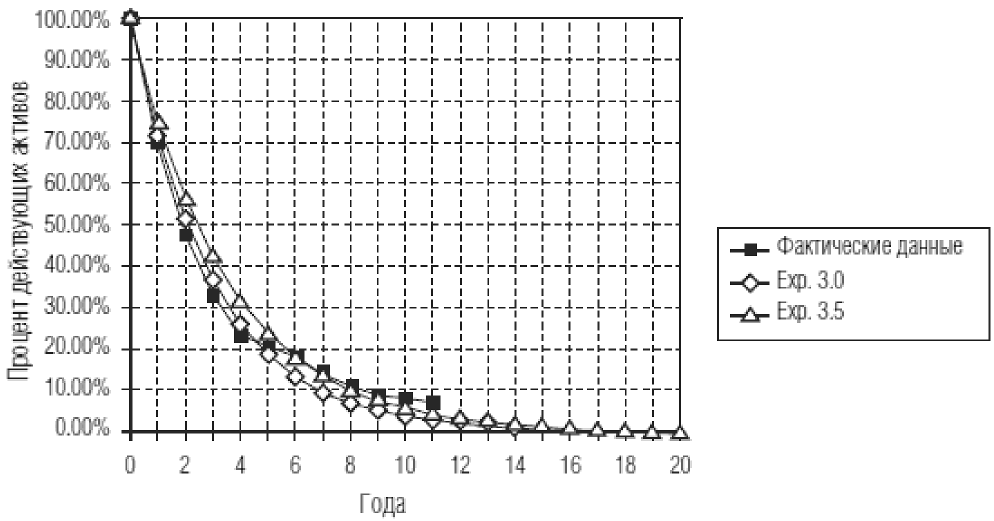


Рис. 5. Пример построения кривой выживаемости. Аппроксимация кривой: экспоненциальная кривая.

Кроме того, для технических систем мы можем рассчитывать оставшееся время эксплуатации (см. статью © *Лейфер Л.А., Кашникова П.М. (выпускница мехмата, специальность «Математичяеские методы экономики»)*, 2007 ЗАО "Приволжский центр финансового консалтинга и оценки").

5. Определение остаточного срока службы машин и оборудования и их остаточной стоимости на основе вероятностных моделей.

Определение остаточного срока службы и остаточного ресурса является важным элементом в процедуре оценки рыночной стоимости машин и оборудования. В рамках затратного подхода остаточный срок службы (остаточный ресурс) необходим для определения остаточной стоимости и соответственно стоимости замещения объекта. При реализации доходного подхода остаточный срок определяет период, в течение которого следует ожидать денежные потоки, и поэтому его величина существенно влияет на расчетную величину рыночной стоимости. При сравнительном подходе остаточный срок службы служит основанием для корректировки цен аналогов, отличающихся от оцениваемого объекта величиной отработавшего времени эксплуатации. Поэтому точность оценки рыночной стоимости машин и оборудования в большой степени зависит от того, насколько правильно определен остаточный срок службы (остаточный ресурс) оцениваемого объекта. В зависимости от того, какой информацией обладает оценщик, возможны различные методы определения остаточного срока службы и остаточного ресурса.

Наиболее надежный прогноз остаточного ресурса может быть осуществлен, если выполнить полномасштабное техническое диагностирование машины с использованием соответствующих средств диагностики. Такой подход требует больших затрат, и поэтому за исключением случаев, когда оцениваются единичные и дорогостоящие машины или технологические линии, в широкой практике оценочной деятельности обычно не применяется. Методы индивидуального прогнозирования остаточного ресурса машин и конструкций, основанные на моделях физических процессов износа машин и конструкций (накопление усталостных повреждений, изнашивание механизмов и т. п.), изложенные в различных публикациях, также не нашли практического применения при оценке стоимости машин в связи с их трудоемкостью и необходимостью применения сложного математического аппарата теории случайных процессов.

Проблема оценки стоимости больших массивов оборудования и машин привела к необходимости создания упрощенных технологий, обеспечивающих оценку «потоком», используя минимум входной информации об объекте оценки. Заметим, что остаточный срок службы (ресурс) в рамках данных моделей обычно определяется как разность между некоторым нормативным сроком службы и его эффективным возрастом.

Рассмотрим статистический подход к задаче прогнозирования остаточного срока службы (ресурса) на основе моделей, которые могут оказаться наиболее приемлемыми во многих реальных ситуациях, связанных с оценкой машин в условиях, когда потеря стоимости в основном обусловлена физической деградацией объекта оценки.

Основные понятия, термины и определения.

Предельное состояние – состояние объекта, при котором его дальнейшая эксплуатация недопустима или нецелесообразна, либо восстановление его работоспособного состояния невозможно или нецелесообразно.

Примечания:

1. Объект может перейти в предельное состояние, оставаясь работоспособным, если, например, его дальнейшее применение по назначению станет недопустимым по требованиям безопасности, экономичности и эффективности.
2. Достижение предельного состояния не сводится только к физическому износу. Как видно из определения, переход в предельное состояние может быть обусловлен также влиянием факторов функционального устаревания.
3. Обычно при достижении предельного состояния объект снимается с эксплуатации. Это, однако, не означает, что стоимость объекта, достигшего предельного состояния, равна нулю. Как показал анализ литературы (и это подтвердилось нашими исследованиями), стоимость объекта, достигшего предельного состояния, обычно составляет 10 – 20% от начальной стоимости. Эта стоимость может включать стоимость оставшихся деталей, материалов и т. п.

Срок службы объекта – календарное время, равное периоду эксплуатации, отсчитываемое от ввода в эксплуатацию объекта до достижения предельного состояния (снятия с эксплуатации).

Ресурс объекта - полная наработка объекта, выраженная в часах, километрах и т. п. отсчитываемая от ввода в эксплуатацию объекта до достижения предельного состояния (снятия с эксплуатации).

Примечания:

1. При стандартной эксплуатации обычно наработка, измеренная в часах или километрах (для транспортных средств), связана пропорциональной зависимостью со сроком эксплуатации. Поэтому в дальнейшем мы не делаем различия между этими понятиями и будем использовать один из этих терминов, понимая, что все формулы, рассуждения и выводы, относящиеся к одному из них, в той же степени относятся и к другому.
2. Фактические моменты достижения объектами предельного состояния могут существенно различаться в зависимости от индивидуальных свойств и условий эксплуатации объектов. Поэтому срок службы, также как и ресурс объекта, следует считать случайными величинами. Они могут описываться только вероятностными моделями. В качестве такой модели обычно используется плотность распределения или закон распределения. В экономической методологии используется близкое понятие: «кривая выживаемости». Более подробно о вероятностных моделях в следующей главе.

Средний срок службы (Средний ресурс) – среднее значение случайной величины - срока службы (ресурса), отсчитываемое от ввода в эксплуатацию объекта до достижения предельного состояния (снятия с эксплуатации).

Установленный (Нормативный) срок службы (установленный ресурс) - срок эксплуатации, установленный в технической документации.

Примечания:

1. Установленный (Нормативный) срок службы характеризует долговечность объекта, его способность в течение установленного срока сохранять эксплуатационные характеристики. Изъятие объекта из эксплуатации по причине достижения предельного состояния до завершения установленного срока эксплуатации считается маловероятным. При этом достижение объектом нормативного срока не означает, что объект достиг предельного состояния и должен быть снят с эксплуатации. Чтобы обеспечить уверенную эксплуатацию объекта в течение установленного срока, объект должен иметь некоторый запас прочности, который дает возможность уверенно эксплуатировать объект в течение нормативного срока и еще некоторое время после окончания этого срока. Проводимые на заводе – изготовителе отработка и испытания объекта направлены на обеспечение надежной эксплуатации в течение установленного срока (установленного ресурса) и на обеспечение этого запаса. С вероятностной точки зрения установленный в документации срок представляет собой квантиль распределения ожидаемого срока службы.
2. Следует различать средний срок службы и нормативный срок службы. Нормативный срок службы не является средним сроком службы, но он может использоваться в качестве исходной информации для определения среднего срока службы и других статистических параметров, характеризующих долговечность объекта.
3. Если в конструкторской или эксплуатационной документации не указан срок эксплуатации, то в качестве нормативного срока может выступать величина, рассчитанная на основе нормы амортизации объекта данного класса. По смыслу такая величина также характеризует долговечность объекта.

Возраст объекта – период времени от даты начала эксплуатации до текущего момента.

Остаточный срок службы – календарная продолжительность эксплуатации от текущего момента до достижения им предельного состояния. Отличается от срока службы тем, что в качестве начала отсчета принимается текущий момент, до которого он уже некоторое время эксплуатировался, а не начало эксплуатации.

Остаточный ресурс объекта - наработка объекта, выраженная в часах, километрах и т. п., от текущего момента до достижения им предельного состояния. Отличается от ресурса объекта тем, что в качестве начала отсчета принимается текущий момент, до которого он уже некоторое время эксплуатировался, и часть начального ресурса исчерпал.

Примечания:
1. Индивидуальные характеристики объекта (остаточный срок службы и остаточный ресурс) являются случайными величинами и точно могут быть определены лишь после того, как наступило его предельное состояние. Пока эти события не наступили, можно лишь говорить о прогнозировании этих величин с большей или меньшей вероятностью. Поэтому остаточный срок службы является прогнозируемым значением ожидаемого времени, по окончании которого объект достигнет предельного состояния и будет снят с эксплуатации. Следует особо подчеркнуть, что остаточный срок в общем случае не равен оставшемуся времени до достижения нормативного срока. Это же относится и к остаточному ресурсу.
2. Поскольку остаточный срок службы (остаточный ресурс) - случайная величина, она может описываться только вероятностными моделями. В качестве такой модели так же, как и в случае начального срока службы (ресурса), может использоваться кривая выживаемости.

Средний остаточный срок службы (Средний остаточный ресурс) - среднее значение случайной величины - остаточного срока службы (ресурса), отсчитываемого от текущего момента до достижения предельного состояния (снятия с эксплуатации).

Примечания:
1. Следует четко понимать, что средний остаточный срок службы не показывает точный период времени, который будет эксплуатироваться оцениваемый объект. Он характеризует некоторый центр рассеивания моментов времени, вокруг которого (часть раньше, часть позже) будут сниматься с эксплуатации объекты данного класса, которые достигли предельного состояния. Поскольку на момент оценки нельзя определить точное время, которое объект еще способен эксплуатироваться, средний остаточный ресурс представляет собой наилучший ориентир для ожидаемого срока службы оцениваемого объекта.
2. Средний остаточный срок службы зависит от начальных характеристик долговечности объекта и его возраста. Чем больше возраст объекта, тем меньше его средний остаточный срок. Таким образом, средний остаточный срок убывает по мере увеличения возраста объекта оценки. Однако достижение нормативного срока не означает, что средний остаточный срок службы равен нулю.

Вероятностные модели для описания срока службы (ресурса)

Поскольку срок службы является случайной величиной, для его описания следует использовать вероятностные модели. Вероятность того, что за время объект не достигнет предельного состояния определяют как $S(t) = P(X > t)$. Функция $S(t)$ показывает: какова вероятность того, что объект проживет больше, чем время t Заданная таким образом кривая выживаемости связана с функцией распределения вероятностей $F(t)$ соотношением $F(t)=1-S(t)$.

При этом, если отсчет времени ведется от текущего момента τ, характеризующего время, до которого объект уже эксплуатировался, то $F(t|\tau)$ характеризует распределение вероятностей случайной величины - остаточного срока службы. На языке теории вероятностей $F(t|\tau)$ - условная вероятность того, что остаточный срок службы будет не менее t при условии, что объект исправно функционировал до текущего момента - τ.

Следует различать теоретическое распределение вероятностей и эмпирическое (или выборочное, т. е. построенное по выборочным данным). Построить эмпирическое распределение на основе статистических данных не представляет принципиальных трудностей. Однако для того, чтобы эмпирическое распределение могло быть непосредственно использовано для установления теоретического распределения, необходимы большие объемы данных. Поэтому все выводы относительно теоретического распределения делаются на основе анализа природы данных, характера процессов, приводящих к предельному состоянию и ограниченного объема выборочных данных.

Для описания срока полезного действия используются известные вероятностные распределения, в частности, модель Вейбулла и модели выживаемости типа Айова. Наряду с моделями, предложенными в штате Айова, для вероятностного описания срока службы машин, механизмов, сложных технических систем может использоваться также логнормальное распределение, которое наряду с распределением Вейбулла получило широкое применение и развитие в теории надежности технических систем, машин и сложных конструкций. Выбор того или иного распределения определяется характером преобладающих физических процессов, наличием исходной информации и возможностями вычислительных процедур.

Здесь рассматривается модель, позволяющая при принятых допущениях ответить на эти вопросы и тем самым создать реальные предпосылки для практического использования вероятностных моделей в задачах определения остаточного срока службы машин и оборудования. В качестве такой модели используется логнормальное распределение, которое в наибольшей степени адекватно процессам физического изнашивания, усталостного накопления повреждений и другим механизмам потери работоспособности машин и механизмов. Логарифмически нормальное распределение можно вывести как статистическую модель для случайной величины, значения которой получаются в результате умножения большого числа случайных факторов. Логарифмически нормальное распределение применяется в самых различных областях – от экономики до биологии для описания процессов, в которых наблюдаемое значение составляет случайную долю предыдущего значения. Обоснование применимости логарифмически нормального распределения для описания срока службы также основано на свойстве умножения эффектов, присущем данному распределению. Поэтому данное распределение получило широкое применение и развитие в работах по анализу процессов деградации механических систем.

Далее будут использоваться безразмерные характеристики календарного времени, равные отношению соответствующей временной характеристики к нормативному сроку службы τ_x.

Обозначим безразмерное время, равное отношению срока службы τ к нормативному сроку службы τ_x, буквой t: $t = \tau / \tau_x$.

Тогда в соответствии с принятой моделью срока службы плотность распределения случайной величины (t) имеет вид:

$$f(t;\mu,\sigma) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln t - \mu)^2}{2\sigma^2} \right), t > 0, -\infty < \mu < \infty, \sigma > 0.$$

Плотность распределения содержит всю информацию относительно срока службы. Однако непосредственно для проведения оценки необходимо знание основных характеристик данного распределения (μ и σ).

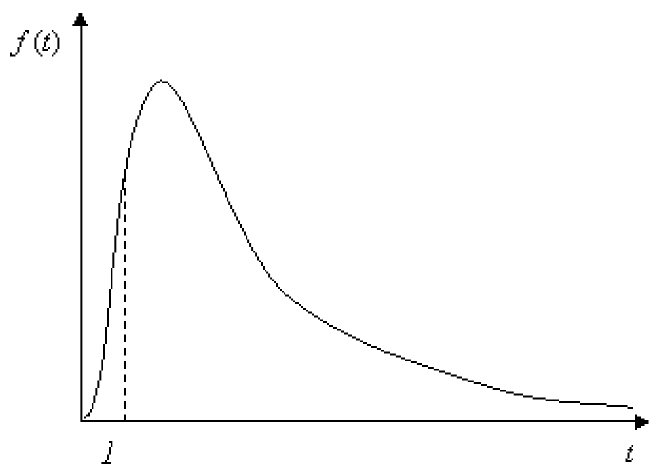


Рис. 6. Плотность распределения случайной величины (*t*).

Математическое ожидание T, дисперсия D и коэффициент вариации ρ случайной величины $t-$ срока службы, заданного в безразмерном виде определяются через параметры распределения μ и σ следующим образом:

$$T = e^{\mu+\sigma^2/2},\ D = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1),\ \rho = \sqrt{D} / T = \sqrt{e^{\sigma^2} - 1}. \qquad (10)$$

Для определения параметров распределения следует воспользоваться информацией, доступной оценщику. В качестве такой информации могут использоваться общие сведения относительно объекта оценки и нормативный срок службы, заданный в эксплуатационной документации. Как отмечалось выше, если отсутствуют данные о сроке службе, можно воспользоваться нормами амортизации, которые также несут информацию об оцениваемом объекте.

Проанализируем информацию, которая позволяет определить основные характеристики логнормального распределения.

Анализ литературы, обобщающей многочисленные исследования по надежности и долговечности машин и оборудования, показывает, что коэффициент вариации для машин и оборудования лежит в пределах: 0.3 – 0.4. Эта информация позволяет определить параметр распределения σ. Для того, чтобы нормативный срок службы, относящийся к данному объекту, можно было использовать для определения параметров распределения, учтем, что нормативный срок службы представляет собой календарное время, в течение которого объект должен исправно функционировать (более точно, не должен достигнуть своего предельного состояния). По существу, нормативный срок службы указывает минимальное время, в течение которого объект должен эксплуатироваться, если не происходит каких – либо нештатных ситуаций. Таким образом, если предположить, что объект с высокой вероятностью (например, 0.9) должен прослужить в течение заданного срока, то с точки зрения принятой модели нормативный срок представляет собой 10–процентный квантиль распределения. Используя указанную выше информацию и соответствующие допущения, легко рассчитать параметры логнормального распределения и построить кривую выживаемости, характеризующую процесс выбытия оцениваемых объектов за период эксплуатации.

Зададим уровень α, он будет представлять собой вероятность того, что объект оценки достигнет предельного состояния до истечения нормативного срока, который в свою очередь определяется интегралом

$$\int_0^1 f(t;\mu,\sigma)\,dt = \alpha \qquad (11)$$

Используя данное уравнение (11) и соотношения (10), можно рассчитать значения безразмерного среднего срока службы T по заданным значениям α и ρ. Напомним, что безразмерный средний срок службы (T) является величиной, равной отношению среднего значения фактического срока службы к нормативному сроку службы.

В табл. 3 представлены результаты таких расчетов для различных значений α и ρ.

Таблица 3. Значения безразмерного среднего срока службы (T)

ρ \ α	0,05	0,1	0,2
0,2	1,419068	1,320486	1,20925
0,3	1,716007	1,540335	1,349859
0,4	2,095936	1,814845	1,521962
0,5	2,58571	2,159766	1,733253

Так же можно рассчитать параметры логнормального распределения, характеризующего вероятностные свойства процесса выбытия объектов оценки из эксплуатации. На рис. 7 представлена кривая выживаемости (иногда ее называют кривой смертности), описывающая процесс выбытия объектов из эксплуатации.

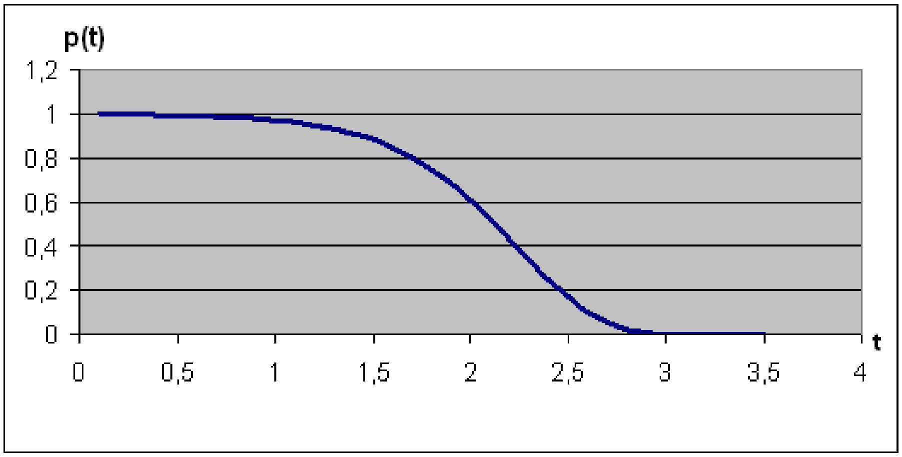


Рис. 7. Кривая выживаемости (ρ=0.3, α=0.1)

При этом кривая выживаемости построена, исходя из условий: ρ=0.3, α=0.1. Основанием для выбора таких исходных данных послужили два обстоятельства:

1. Предельное состояние у механических систем наступает в основном из–за процессов физического изнашивания и усталостного накопления повреждения. Поэтому, опираясь на многочисленные исследования в теории надежности (см., например, [1,4]), в качестве коэффициента вариации может быть принята величина, равная 0.3 – 0.4.
2. Нормативный срок (назначенный), указанный в конструкторской или эксплуатационной документации, представляет собой не что иное, как минимально допускаемый срок эксплуатации объекта, в течение которого объект не должен достигать своего предельного состояния. Поскольку, тем не менее, такую возможность нельзя исключать полностью, мы исходим из того, что объект снимается с эксплуатации и списывается не более, чем в 10% случаев. В результате кривая выживаемости характеризует в основном процесс выбытия объектов в период времени после нормативного срока службы. Естественно, что в соответствии с таким предположением средний срок службы объекта, который используется в дальнейших расчетах по оценке, превышает нормативный срок службы, что вполне оправдано с точки зрения реальной картины рынка.

Остаточный срок службы.

Если объект достиг некоторого возраста, то естественно ожидать, что остаточный срок службы для него несколько уменьшится. При этом, чем выше возраст объекта (при условии одинаковой истории жизни объектов), тем меньше его остаточный срок. Это утверждение отвечает всем известным моделям потери стоимости и здравому смыслу.

В этом случае распределение остаточного срока службы оцениваемого объекта и соответственно кривая выживаемости, характеризующая вероятностный процесс выбытия объектов данного класса доживших до данного возраста, могут быть рассчитаны, исходя из условного распределения вероятностей. Условная плотность логарифмически нормального распределения остаточного срока службы, выраженного в относительных единицах, отвечающая условию, что объект дожил до возраста τ, определяется следующим образом:

$$f(t) = \begin{cases} \dfrac{\exp\left(-\dfrac{(\ln t - \mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma t}, & t > \dfrac{\tau}{\tau_n}, \\ 0, & t \le \dfrac{\tau}{\tau_n}. \end{cases}$$

Дальнейшие расчеты и соответствующие графики построены в предположении, что коэффициент вариации ρ=0,3 и допустимый уровень выбытия объектов из эксплуатации до достижения ими нормативного срока α=0,1

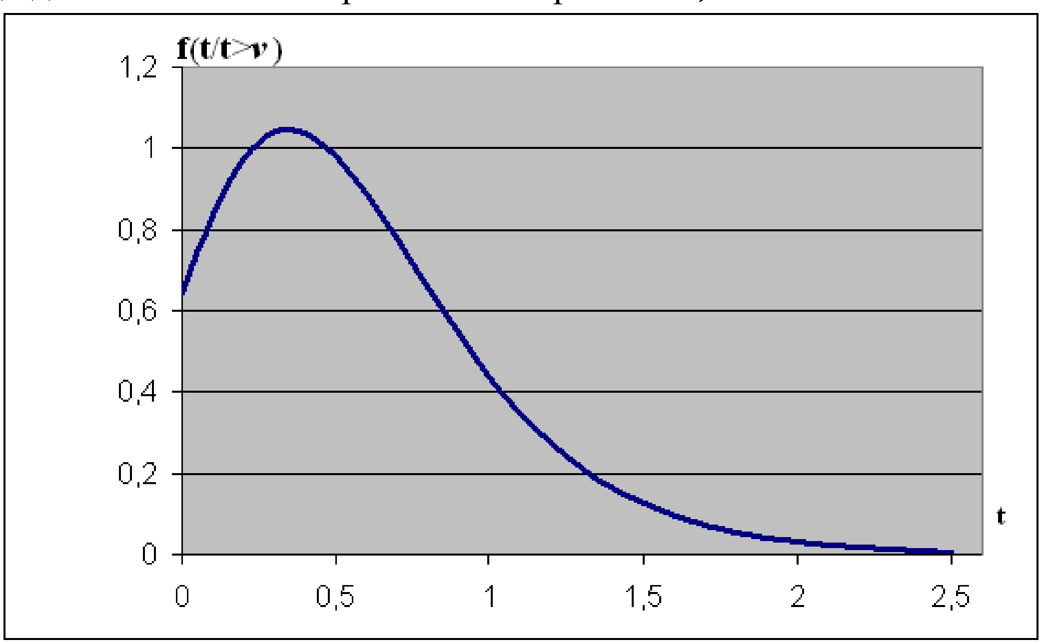


Рис. 8. Условная плотность распределения остаточного срока службы при условии, что объект эксплуатировался до текущего момента.

Заметим, что ν - возраст объекта на момент оценки в относительных единицах, численно равный фактическому времени эксплуатации, деленному на нормативный срок службы: $\nu = \tau / \tau_н$.

Зная плотность распределения остаточного срока службы, можно определить среднее значение остаточного срока службы T (в относительных единицах) при условии, что объект уже эксплуатировался некоторое время (τ).

Ниже приводится зависимость среднего значения остаточного срока службы от фактического срока эксплуатации, предшествующего дате оценки. Эта зависимость построена путем статистического моделирования случайных величин, генерируемых упомянутой плотностью распределения, и последующего расчета среднего значения и медианы. Полученные результаты отражают вероятностную природу долговечности машин и более соответствуют реалиям, чем детерминированные модели. В частности, они учитывают, что достижение объектом нормативного срока не означает, что ресурс полностью исчерпан.

При параметрах, заложенных в приведенных расчетах, объект, отработавший свой нормативный срок, сохраняет возможность дальнейшей эксплуатации в среднем еще в течение времени до 40% от нормативного срока. Оставшийся срок учитывает заложенный запас по ресурсу машины, поскольку нормативный срок не есть срок полного исчерпания ресурса. Из графика также видно, что с увеличением предшествующего срока эксплуатации среднее значение остаточного срока службы убывает, и объект, проработавший существенно больше своего нормативного срока службы, ожидает в скором времени достижение предельного состояния.

Приведенные ниже примеры, показывают, как изложенной теорией можно пользоваться в практических расчетах в процессе оценки рыночной стоимости машин и оборудования.

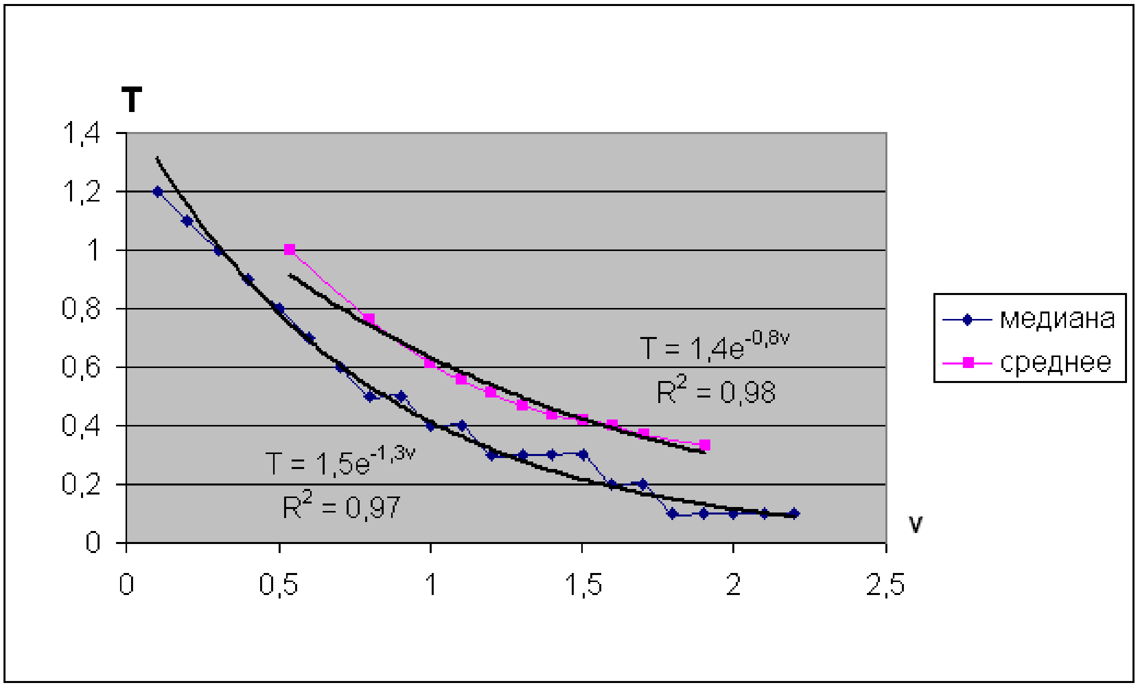


Рис. 9. Зависимость среднего значения остаточного срока (*T*) от предшествующего срока эксплуатации (*v*).

Примеры расчета остаточного срока движимого имущества.

В заключение приведем примеры определения среднего остаточного ресурса, иллюстрирующие процесс оценки остаточного срока службы при оценке машин и оборудования с использованием графика для среднего значения остаточного ресурса (рис. 9).

Пример 1.

1. Объектом оценки является сложная технологическая линия с заданным нормативным сроком службы, равным 20 лет.
2. Оборудование приобретено у диллеров и поставлено на эксплуатацию 14 лет тому назад. Линия эксплуатировалась в штатных условиях с соблюдением всех требований эксплуатационной документации (планово – предупредительные ремонты, профилактика и пр.) В настоящее время она находится в рабочем состоянии.
3. Деградационные процессы происходили под воздействием физического изнашивания и усталостного накопления повреждений. Коэффициент вариации поэтому можно принять равным 0.3.
4. При определении среднего остаточного срока службы требуется, чтобы был установлен период, в течение которого следует ожидать, что объект будет генерировать денежные потоки. Данная величина требуется для реализации доходного подхода.

Расчет

В качестве исходных данных используются:
нормативный срок – 20 лет,
текущий возраст – 14 лет (в относительных единицах 14/20 = 0.7).
Из графика определяем средний остаточный срок службы в относительных единицах, который составит 0,6. Отсюда средний остаточный срок – 0,6 * 20 = 12 лет.

Пример 2.

1. Объект оценки – сельскохозяйственный трактор, нормативный срок службы согласно конструкторской документации – 12 лет
2. Трактор был приобретен в торговой сети и эксплуатировался в штатном режиме полный срок службы - 12 лет.
3. На текущий момент трактор является работоспособным, т. е. способным выполнять заданные функции в соответствии с требованиями нормативно – технической и конструкторской документации. Ресурсные параметры находятся в допустимых пределах.
4. Деградационные процессы, относящиеся к ресурсным параметрам (зазоры в сопряжениях, износы в подшипниках, шестеренках, валах и т. п.), происходили в основном под воздействием физического изнашивания. Поэтому коэффициент вариации срока службы можно принять равным 0,3.
5. При определении остаточного срока службы требуется определить величину потери стоимости объекта, прослужившего полный срок службы и не достигшего при этом предельного состояния, в рамках затратного подхода.

Расчет

Исходные данные:
нормативный срок – 12 лет,
текущий возраст – 12 лет (в относительных единицах 12/12 = 1).
Из графика определяем средний остаточный срок службы в относительных единицах: 0.4.
Таким образом, средний остаточный срок: 0.4* 12 = 4.8 лет.
Отсюда, если считать величину износа методом экономической жизни, получим: Износ = текущий возраст/(текущий возраст + средний остаточный ресурс). Износ = 12/ (12+4.8) = 0.7. Используя полученную величину износа в качестве исходных данных, можно рассчитать текущую стоимость объекта.

Пример 3.

1. Объект оценки - импортный легковой автомобиль выпуска 1993 года, приобретен на вторичном рынке. На текущий момент возраст автомобиля – 11 лет.
2. В эксплуатационной документации отсутствует нормативный срок эксплуатации. Однако некоторое представление о нем дают нормы амортизации, отражающие средний срок эксплуатации объектов данного класса. Исходя из норм амортизации, нормативный срок эксплуатации автомобиля данного класса - 7 лет.

3. На текущий момент автомобиль является работоспособным, т. е. способным выполнять заданные функции в соответствии с требованиями нормативно – технической и конструкторской документации. Ресурсные параметры находятся в допустимых пределах.
4. Деградационные процессы, относящиеся к ресурсным параметрам (зазоры в сопряжениях, износы в подшипниках, шестеренках, валах и т. п.), происходили в основном под воздействием физического изнашивания. Поэтому коэффициент вариации срока службы можно принять равным 0,3.
5. Несмотря на то, что автомобиль отслужил нормативный срок эксплуатации, поскольку автомобиль находится в хорошем состоянии, принято решение продолжать его эксплуатацию. Это должно найти отражение в оценке рыночной стоимости основных средств предприятия. Для этого требуется определить остаточный срок эксплуатации.

Расчет

Используем в качестве исходных данных:
нормативный срок – 7 лет,
текущий возраст – 11 лет (в относительных единицах 11/7 = 1,5).
Из графика определяем средний остаточный срок службы (в относительных единицах): - 0,3
Таким образом, средний остаточный срок – 0,3* 7 = 2,1 года.

6. Список использованной литературы.

1. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика: Начальный курс. – М., Дело, 2004 – 576 с.
2. Greene W.H. Econometric Analysis, 5rd edition. – NY, Prentice-Hall, 2002. – 958 p.
3. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика. – М., Экзамен, 2003. – 512 с.
4. Дорохина Е.Ю., Преснякова Л.Ф., Тихомиров Н.П. Сборник задач по эконометрике. – М., Экзамен, 2003. – 224 с.
5. Катышев П.К., Магнус Я.Р., Пересецкий А.А. Сборник задач к начальному курсу эконометрики. – М., Дело, 2003. – 208 с.
6. Тихов М.С. Эконометрические дискретные модели бинарного выбора. – Прикладная статистика в социально-экономических проблемах: материалы межд. конф., т.1, изд-во ННГУ, Нижний Новгород, 2003, с.104-106.
7. Kaplan, E. L.; Meier, P.: Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, 1958, v.**53**, p.457–481.
8. Aalen O.O. Nonparametric inference for a family of counting process. – Ann. Statist., 1978, v.6, p.701-726.
9. Abdushukurov A.A. Estimation of the probability density and intensity function of the Koziol-Green model of random censoring. – Sankhya, Ser.A, 1987, v.4, p.150-168.
10. Nelson W. Theory and applications for hazard plotting for censored failure data. – Technometrics, 1972, v.14, p.945-965.
11. Ageev V.V., Tikhov M.S. Estimating of the survival function with doubly censored data. – Proceedings of the 9th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’09), 21–24 October 2009, Riga, Latvia, p. 134-143. ISBN 978-9984-818-21-4, Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia, p. 134-143. http://www.tsi.lv/?id=1559&lang=en&ct=8&r=6&top=0
12. Тихов М.С., Агеев В.В., Бородина Т.С. Оценивание параметров распределения Вейбулла по случайно цензурированным выборкам. – ж. «Вестник Нижегородского университета им. Н.И. Лобачевского», Н.Новгород, изд-во ННГУ, №4, 2010, с.141-145.
13. Рейли Р., Швайс Р. Оценка нематериальных активов. М.: ИД «Квинто-консалтинг», 2005 – 772 с.
14. Лейфер Л.А., Кашникова П.М., 2007 ЗАО "Приволжский центр финансового консалтинга и оценки" Определение остаточного срока службы машин и оборудования на основе вероятностных моделей Режим доступа: http://www.labrate.ru/leifer/leifer_kashnikova_article_2007-1_residual_service_life.htm
15. Тихов М.С. О сокращении длительности испытаний при цензурировании выборки. ж. «*Теория вероятностей и её применения*», 1991, т.36, в.3, с.626-629.

Содержание.