



Методы статистического анализа в задачах планирования закупок запасных частей серверного оборудования

Студент: Дзахоев Альберт Таймуразович

Руководитель ОИЯИ: Балашов Никита

Александрович

Руководитель СОГУ: Олисаев Эльбрус Георгиевич

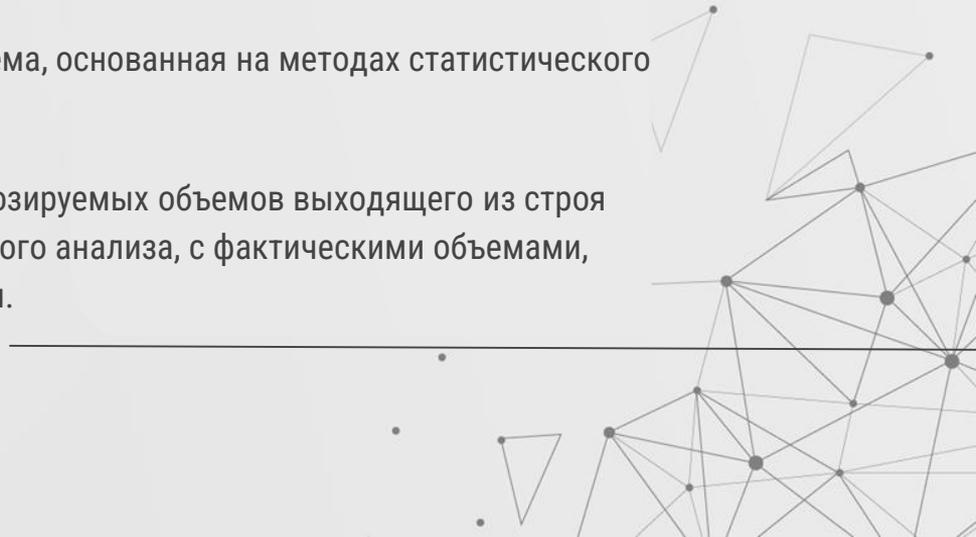
Цель и постановка задачи

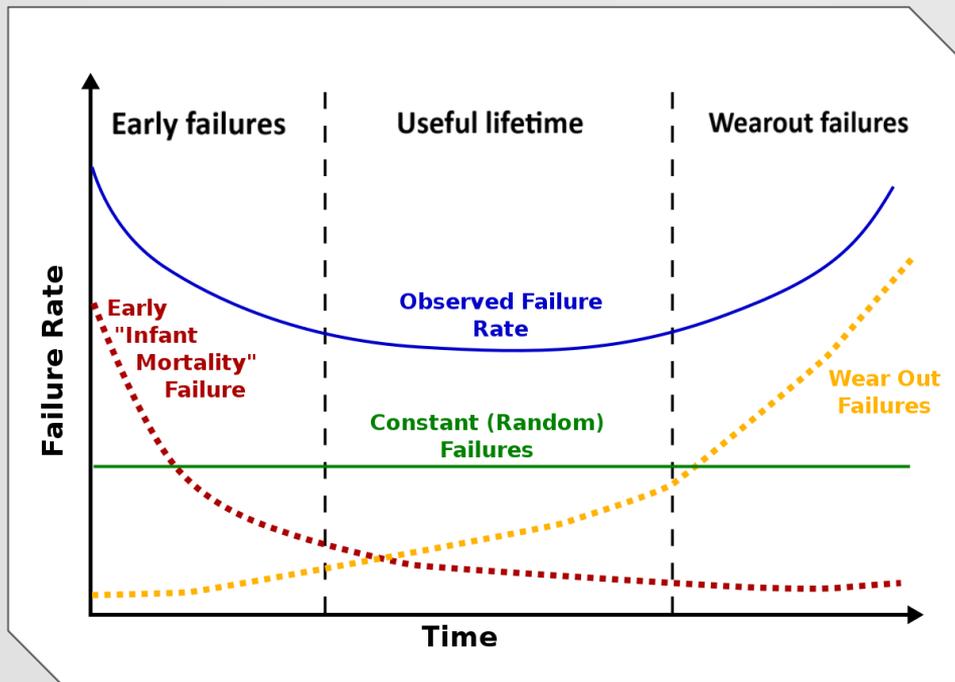
Актуальность: данная тема актуальна вследствие изменений в сфере гарантийного обслуживания.

Цель работы: Разработка программного инструмента для проведения статистического анализа и прогнозирования объемов выходящего из строя оборудования.

Результат: Разработанная рекомендательная система, основанная на методах статистического анализа.

Критерии оценки результата: Сопоставление прогнозируемых объемов выходящего из строя оборудования, полученных на основе статистического анализа, с фактическими объемами, наблюдаемыми в реальных условиях эксплуатации.





Bathtub curve

bathtub curve - это график, отражающий изменение частоты отказов компонентов со временем. Он состоит из трех основных областей: **Early failures** с высокой частотой отказов, **Useful lifetime** с близкой к постоянной частотой отказов из-за случайных событий и **Wearout failures** с возрастающей частотой отказов из-за износа.

Задачи

01

Собрать данные для проведения пробного анализа

02

Изучить методы статистического анализа

03

Организовать программную исследовательскую среду

04

Создать программную реализацию выбранных методов

05

Встроить разработанные инструменты в систему инвентаризации



```
=== START OF INFORMATION SECTION ===
Vendor:                SEAGATE
Product:               ST18000NM004J
Revision:              E004
Compliance:           SPC-5
User Capacity:        18,000,207,937,536 bytes [18.0 TB]
Logical block size:   512 bytes
Physical block size:  4096 bytes
LU is fully provisioned
Rotation Rate:        7200 rpm
Form Factor:          3.5 inches
Logical Unit id:      0x5000c500dad5185f
Serial number:        ZR5ETNBY0000W332TVTX
Device type:          disk
Transport protocol:   SAS (SPL-3)
Local Time is:        Mon Oct 23 12:35:22 2023 MSK
SMART support is:     Available - device has SMART capability.
SMART support is:     Enabled
Temperature Warning:  Enabled
Read Cache is:        Enabled
Writeback Cache is:   Enabled

=== START OF READ SMART DATA SECTION ===
SMART Health Status:  OK
```

ОТЧЕТЫ S.M.A.R.T.

S.M.A.R.T. (Self-Monitoring, Analysis, and Reporting Technology) - это технология, используемая в жестких дисках для мониторинга и оценки их состояния. SMART-отчеты представляют собой данные о состоянии жесткого диска, такие как температура, количества ошибок чтения/записи, количество перезагрузок и так далее..

Backblaze

Американская компания, специализирующаяся на предоставлении услуг онлайн-хранилища данных и резервного копирования. Они публикуют подробную статистику о работе тысяч жестких дисков, использованных в их центрах обработки данных

Drive Stats 2023 Snapshot

Drive Count

269,756

Drive Failures

16,962

Drive Days

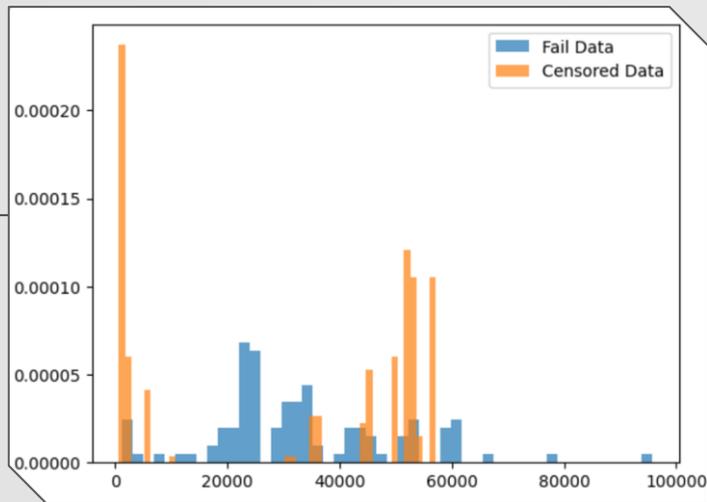
424,474,539

Annualized Failure Rate

1.46%

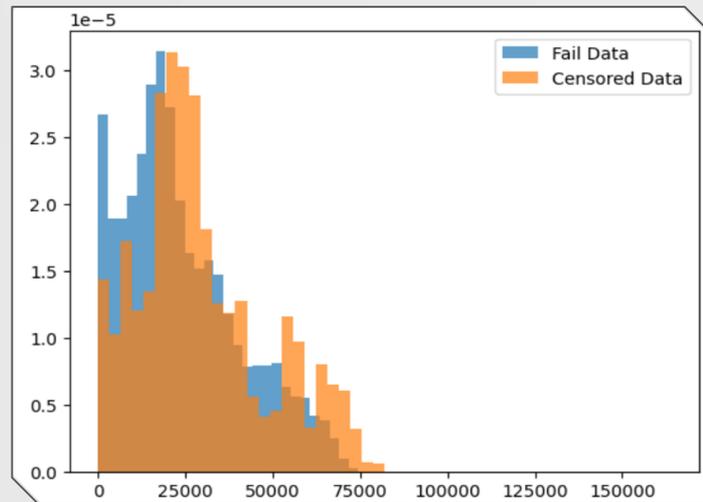
Hard Drive Reliability - Annualized Hard Drive Failure Rates





Гистограмма данных предоставленных ЛИТ

- Fail Data - наработка вышедших из строя дисков
- Censored Data - наработка продолжающих работать дисков



Гистограмма данных Backblaze

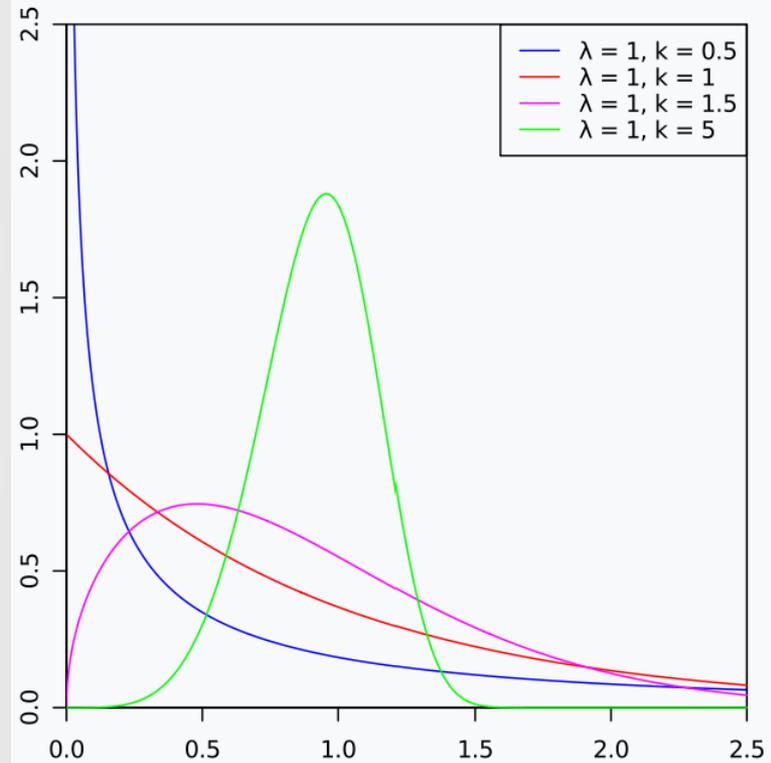
Распределение Вейбулла

Это распределение используется при анализе надежности, например для вычисления среднего времени наработки на отказ какого-либо устройства.

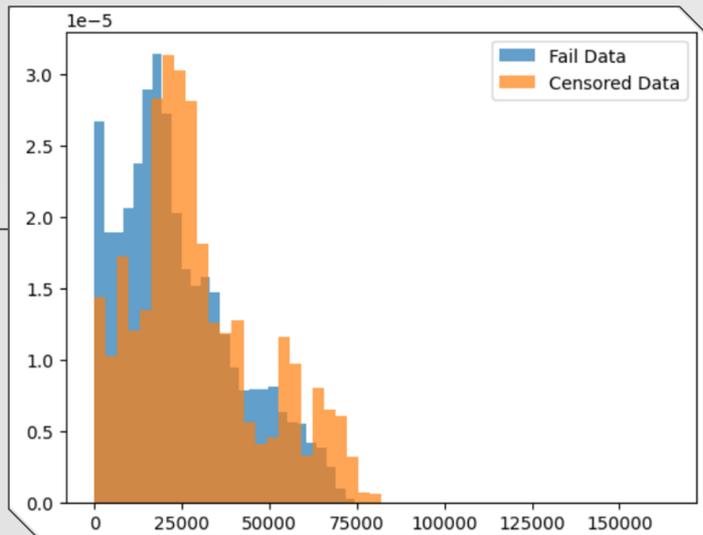
Применяется:

- Моделирование времени между событиями.
- Для моделирования выживаемости.

$$f_X(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



Функция плотности вероятности



Гистограмма используемых данных

- Fail Data - наработка вышедших из строя дисков
- Censored Data - наработка продолжающих работать дисков

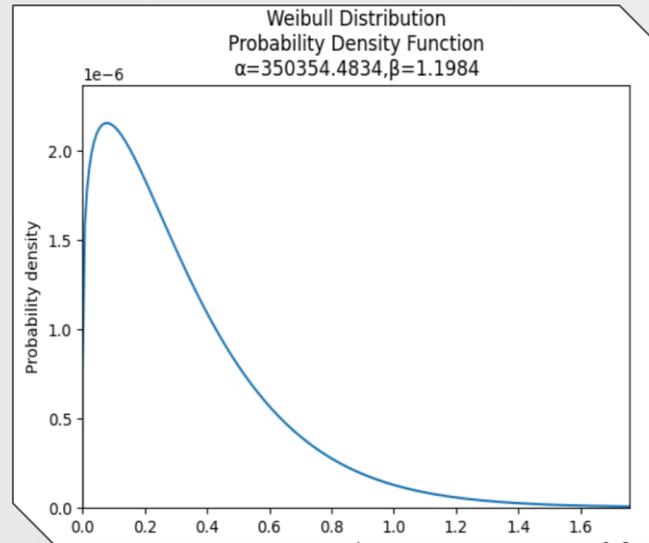


График функции плотности вероятности(PDF)

Функция плотности вероятности (PDF) используется для описания вероятности выхода из строя компонентов в конкретный момент времени или за конкретный период времени.

Google Colaboratory



RELIABILITY

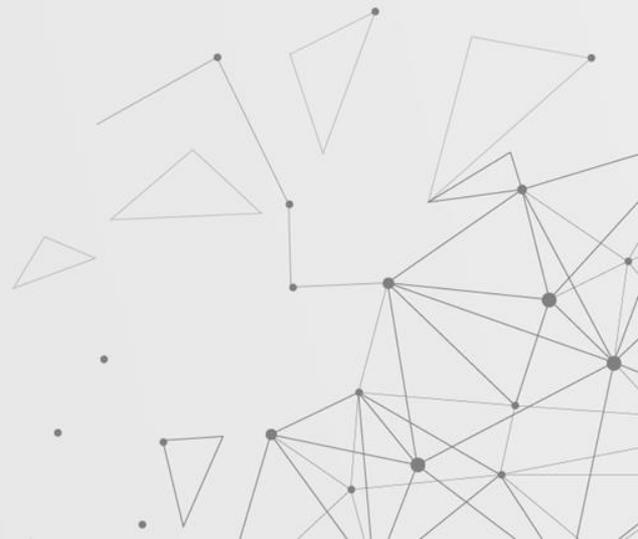
A Python library for reliability engineering

PyPI package v0.8.16 Docs passing Build and Test passing CodeQL passing Code lines 62k
PyPI downloads/month 56k License LGPLv3 DOI 10.5281/zenodo.3937999 Support this project
Provide feedback

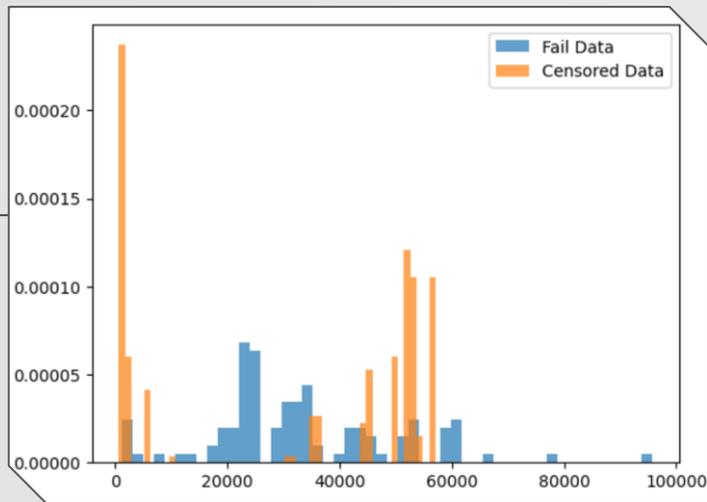
reliability is a Python library for [reliability engineering](#) and [survival analysis](#). It significantly extends the functionality of `scipy.stats` and also includes many specialist tools that are otherwise only available in proprietary software.

RELIABILITY

Библиотека, в которой реализованы различные методы для расчета и анализа данных о надежности, включая распределение Вейбулла.



Результаты анализа



Гистограмма используемых данных

- Fail Data - наработка вышедших из строя дисков
- Censored Data - наработка продолжающих работать дисков

Результаты

Предположительное количество дисков, которые выйдут из строя на промежутке времени с 10 тыс до 20 тыс часов эксплуатации

Results from Fit Weibull 2P (95% CI):

Analysis method: Maximum Likelihood Estimation (MLE)

Optimizer: L-BFGS-B

Failures / Right censored: 113/235 (67.52874% right censored)

| Parameter | Point Estimate | Standard Error | Lower CI | Upper CI |
|-----------|----------------|----------------|----------|----------|
| Alpha | 70680.6 | 4292.38 | 62749.1 | 79614.7 |
| Beta | 1.79142 | 0.150356 | 1.51969 | 2.11173 |

Вероятность выхода из строя до момента времени 10000: 0.029650344090595544

Вероятность выхода из строя до момента времени 20000: 0.098944495194869

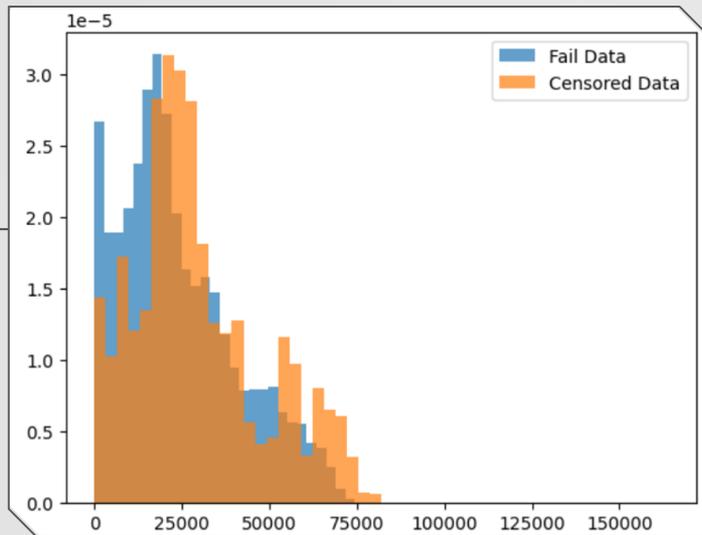
Вероятность выхода из строя до момента времени 30000: 0.19379175535122736

Вероятность выхода из строя до момента времени 200000: 0.998411872986895

Приблизительное количество компонентов,

которые выйдут из строя в промежутке времени [10000, 20000]: 17.323537776068367

Результаты анализа



Гистограмма используемых данных

- Fail Data - наработка вышедших из строя дисков
- Censored Data - наработка продолжающих работать дисков

Результаты

Предположительное количество дисков, которые выйдут из строя на промежутке времени с 10 тыс до 20 тыс часов эксплуатации

Results from Fit Weibull 2P (95% CI):

Analysis method: Maximum Likelihood Estimation (MLE)

Optimizer: TNC

Failures / Right censored: 20897/370412 (94.65972% right censored)

| Parameter | Point Estimate | Standard Error | Lower CI | Upper CI |
|-----------|----------------|----------------|----------|----------|
| Alpha | 350354 | 4995.48 | 340699 | 360284 |
| Beta | 1.19836 | 0.00702269 | 1.18468 | 1.21221 |

Приблизительное количество компонентов,

которые выйдут из строя в промежутке времени [10000, 20000]: 3566.7075093

Задачи

01

Собрать данные для проведения пробного анализа

02

Изучить методы статистического анализа

03

Организовать программную исследовательскую среду

04

Создать программную реализацию выбранных методов

05

Встроить разработанные инструменты в систему инвентаризации