Sprawozdanie

Laboratorium 2

Zadanie 3

GRUPA WCY21IJ1N1

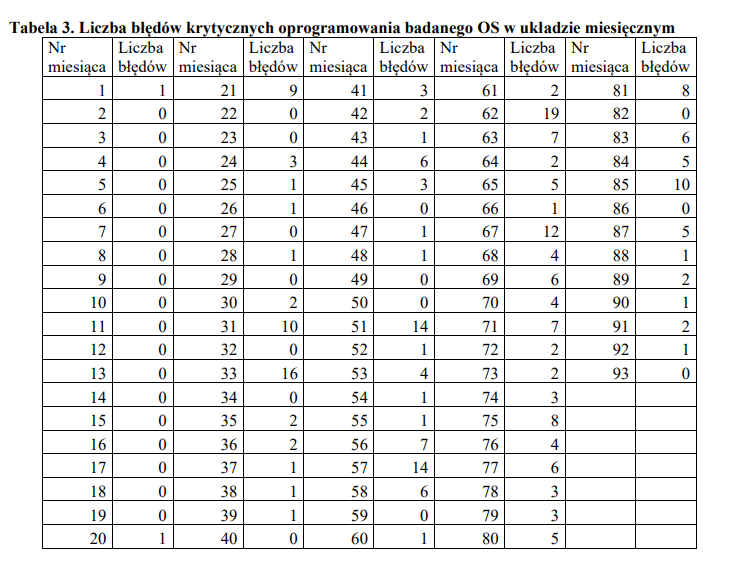
Opracowali:

Karczewski Paweł

Tarkowski Adam

# Treść zadania

Przez cały okres eksploatacji pewnego systemu operacyjnego (OS) zbierano dane dotyczące liczby błędów krytycznych wykrytych w tym czasie w oprogramowaniu. Zebrane obserwacje w układzie miesięcznym przedstawiono w tabeli 3.

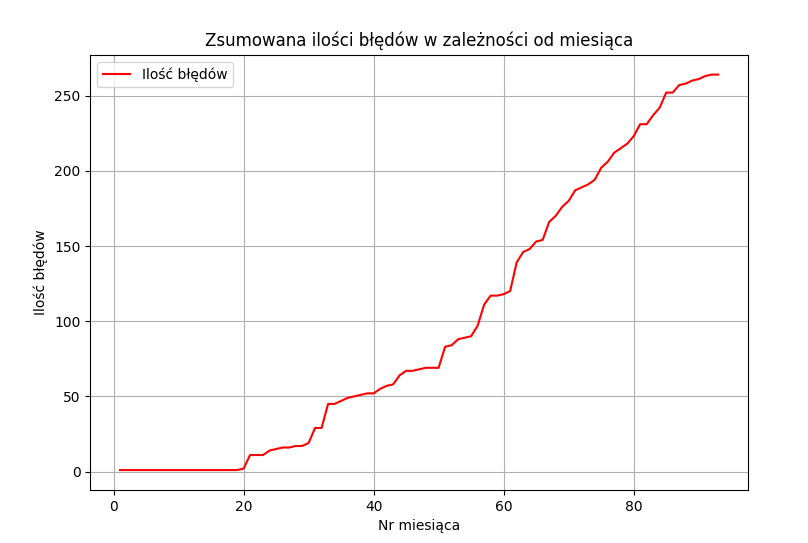


W zadaniu proszę:

1. Wyznaczyć zależność sumarycznej liczby błędów w okresie eksploatacji oprogramowania badanego systemu operacyjnego.
2. Na podstawie opracowanego modelu i przeprowadzonych obliczeń sformułować własne wnioski.
3. Wyniki analizy proszę zawrzeć w postaci sprawozdania, do którego proszę dodać jako załączniki wszystkie pliki z obliczeniami (obliczenia można przeprowadzić w dowolnie wybranym narzędziu)

# Zsumowane błędy w czasie

Na początku naszego zadania trzeba stworzyć dane odnośnie skumulowanej ilości błędów. Po tej operacji wykres wygląda tak:



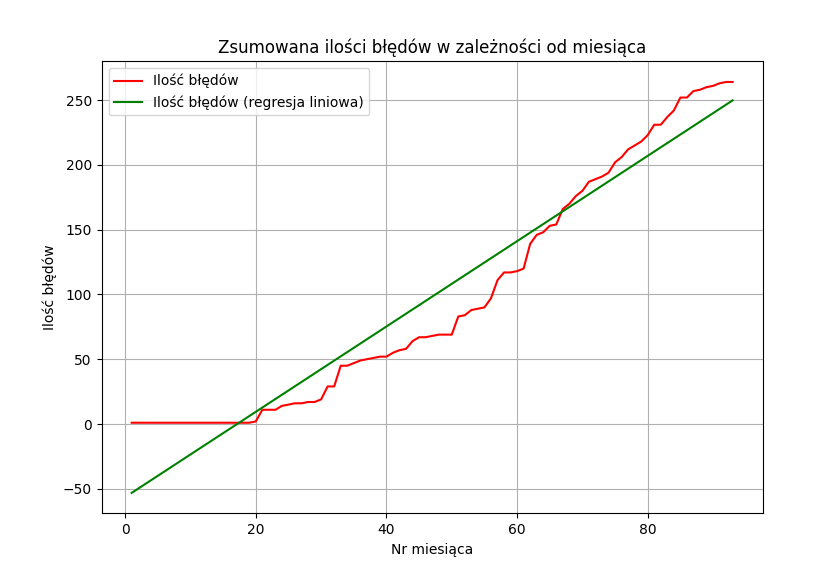
Jak widać wydaje się, że 2 modele które mogą dobrze przybliżać te dane to model liniowy oraz logistyczny.

# Konstrukcja modeli

Na podstawie wylosowanych danych testowych stwórzmy 2 modele.

## Liniowy

Liniowy jest stworzony używając metody najmniejszych kwadratów. Jego wykres wyglądałby w taki sposób:



Błąd średniokwadratowy z tego modelu wynosi ok 603

## Logistyczny

Logistyczny będziemy reprezentować takim wzorem:

Gdzie:

y – to sumaryczna ilość błędów w miesiącu t

a,b,k – parametry funkcji logistycznej

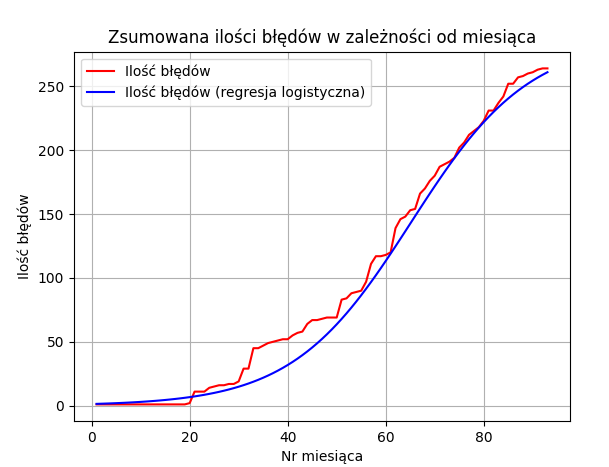
Aby wyliczyć a, b i k skorzystamy z metody Hotellinga. Na początku korzystając z takiego równania wyliczymy parametry a oraz k:

Gdzie:

- przyrost błędów w danym miesiącu

Parametr b liczymy używając tego rówania:

Po wyliczeniu współczynników możemy przetestować jak bardzo mamy przybliżony model:



Błąd średniokwadratowy z tego modelu wynosi ok 130

# Logistyczny przy pomocy optymalizacji

Rozwiążmy zadanie optymalizacyjne żeby wyznaczyć parametry tzn. szukamy takich:

Dla których:

Aby wyznaczyć optymalne parametry skorzystać możemy z metody największego spadku opartej o gradienty, która została zastosowana przy obliczeniach ograniczając ją ilością iteracji.

Zestawienie z regresją liniową:

A graph with a green line and red line

Description automatically generated

A graph with a line

Description automatically generated

Błąd średniokwadratowy dla regresji logistycznej: 99.60069725554209

# Podsumownaie

Jak widać model regresji logistycznej lepiej się spisuje w przybliżaniu wartości dla takiego zestawu danych. Jednym z głównych powodów jest początek danych, gdzie błędy jeszcze nie zaczęły narastać, a możliwe, że jakbyśmy mieli dane z późniejszych okresów to ilość błędów mogłaby się nasycić tzn od pewnego momentu nie zwiększać swojej wartości albo robić to bardzo powoli.