

Eksperymentalny system obsługi zleceń oparty na strumieniowym przetwarzaniu danych

Krzysztof Gogól

Październik 2009

Strumieniowe bazy danych

Aplikacje strumieniowe

Prototypowe DSMS

Problemy otwarte

Model systemu giełdowego

WARSET

Zasady funkcjonowania GPW

Aplikacja

Testy porównawcze

Elementy statystyki

Analiza pojedynczych czynników

Analiza regresji wielokrotnej

Testy Wilcoxon

Podsumowanie

Podział strumieni danych

1. Transakcyjne strumienie danych
 - informacje o użytkownikach stron internetowych (gromadzone w celu personalizacji ustawień)
 - informacje o wykorzystaniu kart bankomatowych (zbierane w celu wykrycia prób defraudacji)
2. Pomiarowe strumienie danych
 - monitorowanie żołnierzy podczas bitwy, natężenia ruchu, temperatury, etc
 - śledzenie natężenie ruchu w Internecie (diagnozowanie ataków)

Najważniejsze cechy aplikacji strumieniowych

1. Brak możliwości wielokrotnego przeglądania strumienia informacji
2. Zmienność danych w czasie
3. Rozproszenie źródeł informacji
4. Niepewność danych
5. Automatyczna reakcja
6. "Ciągłość" w wykonywaniu kwerend
7. Nowy atrybut - czas

Timestamp

- ▶ **Implicit timestamp**
Numer dodawany do każdej krotki informacji przybywającej w strumieniu przez system zarządzania bazą. Może to być numer sekwencyjny oznaczający kolejno odczytywane krotki.
- ▶ **Explicit timestamp**
Własność strumienia informacji. W tym przypadku atrybut *timestamp* oznacza rzeczywisty czas odczytania informacji.

Okna

- ▶ Okna fizyczne
Podział danych oparty na czasie ich przychodzenia.
- ▶ Okna logiczne
Okno zawiera z góry określoną liczbę krotek.
- ▶ *Sliding windows*
Rozmiar okna pozostaje stały, zmianie ulega początek i koniec okna.
- ▶ *Landmark windows*
Okna, w których jeden koniec pozostaje ustalony, a drugi przesuwa się.

Ograniczenia implementacyjne

- ▶ Bezstanowość
Brak możliwości implementacji operacji stanowych uniemożliwia np. wykonanie rzutowania lub selekcji z eliminacją duplikatów, czy grupowania danych.
- ▶ Brak blokad
Brak blokad uniemożliwia prostą implementację operacji agregujących.
- ▶ Działanie w czasie rzeczywistym
Ze względu na konieczność ciągłego działania, automatycznej reakcji i ilości przetwarzanych danych nie jest możliwe wielokrotne przeglądanie danych.

Mierniki jakości DSMS

- ▶ Czas odpowiedzi
Jest to czas oczekiwania na wykonanie kwerendy.
- ▶ Dokładność
W przypadku, gdy w strumieniu informacji znajduje się zbyt wiele danych, których system nie jest w stanie przetworzyć, dopuszczamy odpowiedzi przybliżone. Wówczas można zapytać o dokładność obliczeń na zadanym strumieniu danych.
- ▶ Skalowalność
Skalowalność określa jak wiele zasobów potrzebuje system, by z zadaną dokładnością i zadanym czasem odpowiedzi wykonać daną kwerendę na określonym strumieniu informacji.

Prototypowe DSMS

1. Aurora (2002)
2. COUGAR
3. Gigascope
4. Hancock
5. NiagaraCQ
6. StatStream
7. STREAM
8. TelegraphCQ
9. Tapestry

Problemy otwarte

1. Klastrowanie w strumieniu danych
2. Klasyfikacja w strumieniu danych
3. Poszukiwanie wzorca w strumieniu danych
4. Wykrywanie zmiany w strumieniu danych
5. Obliczenia na danych strumieniowych z wykorzystaniem techniki przesuwających się okien
6. Indeksowanie w strumieniach danych

Strumieniowe bazy danych

- Aplikacje strumieniowe
- Prototypowe DSMS
- Problemy otwarte

Model systemu giełdowego

- WARSET
- Zasady funkcjonowania GPW
- Aplikacja

Testy porównawcze

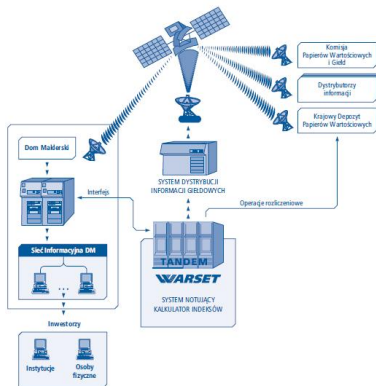
- Elementy statystyki
- Analiza pojedynczych czynników
- Analiza regresji wielokrotnej
- Testy Wilcoxona
- Podsumowanie

WARsaw Stock Exchange Trading System

- ▶ system informatyczny, obsługujący Giełdę Papierów Wartościowych w Warszawie,
- ▶ opracowany przez firmę Hewlett-Packard i działa nieprzerwanie od 17 listopada 2000 roku,
- ▶ początkowo przetwarzał maksymalnie 40 zleceń na sekundę, od czerwca 2007, dzięki wymianie głównego komputera - Tandem firmy Compaq na HP Integrity NonStop NS16000, limit ten został zwiększony do 180 zleceń na sekundę,
- ▶ podobny sprzęt i oprogramowanie jest wykorzystywane przez inne giełdy na świecie (m.in. w Nowym Jorku, Toronto, Paryżu, Brukseli, Amsterdamie).

WARsaw Stock Exchange Trading System

Architektura systemu WARSET



Rysunek: Schemat systemu WARSET

Zasady funkcjonowania GPW

Giełda jest miejscem, gdzie spotyka się podaż z popytem, czyli sprzedający z kupującym.

Na giełdach papierów wartościowych handluje się:

Inwestorzy przekazują zlecenia kupna lub sprzedaży instrumentów finansowych do giełdy papierów wartościowych za pośrednictwem biur maklerskich.

Zasady funkcjonowania GPW

Giełda jest miejscem, gdzie spotyka się podaż z popytem, czyli sprzedający z kupującym.

Na giełdach papierów wartościowych handluje się:

- ▶ akcjami,

Inwestorzy przekazują zlecenia kupna lub sprzedaży instrumentów finansowych do giełdy papierów wartościowych za pośrednictwem biur maklerskich.

Zasady funkcjonowania GPW

Giełda jest miejscem, gdzie spotyka się podaż z popytem, czyli sprzedający z kupującym.

Na giełdach papierów wartościowych handluje się:

- ▶ akcjami,
- ▶ obligacjami,

Inwestorzy przekazują zlecenia kupna lub sprzedaży instrumentów finansowych do giełdy papierów wartościowych za pośrednictwem biur maklerskich.

Zasady funkcjonowania GPW

Giełda jest miejscem, gdzie spotyka się podaż z popytem, czyli sprzedający z kupującym.

Na giełdach papierów wartościowych handluje się:

- ▶ akcjami,
- ▶ obligacjami,
- ▶ kontraktami terminowymi,

Inwestorzy przekazują zlecenia kupna lub sprzedaży instrumentów finansowych do giełdy papierów wartościowych za pośrednictwem biur maklerskich.

Zasady funkcjonowania GPW

Giełda jest miejscem, gdzie spotyka się podaż z popytem, czyli sprzedający z kupującym.

Na giełdach papierów wartościowych handluje się:

- ▶ akcjami,
- ▶ obligacjami,
- ▶ kontraktami terminowymi,
- ▶ opcjami.

Inwestorzy przekazują zlecenia kupna lub sprzedaży instrumentów finansowych do giełdy papierów wartościowych za pośrednictwem biur maklerskich.

Systemy notowań

Na GPW instrumenty finansowe mogą być notowane na jednym z dwóch systemów notowań:

1. systemie notowań jednolitych
2. systemie notowań ciągłych

Decyzje o tym, w jakim systemie są notowane poszczególne instrumenty finansowe podejmuje Zarząd Giełdy na podstawie analizy płynności danego instrumentu.

Priorytety ceny i czasu

Wszystkie zlecenia są realizowane przy zachowaniu dwóch priorytetów:

1. Priorytet ceny

- ▶ Zlecenia kupna z wyższym kursem realizowane są przed zleceniami kupna z niższym kursem.
- ▶ Zlecenia sprzedaży z niższym kursem realizowane są przed zleceniami sprzedaży z wyższym kursem.

2. Priorytet czasu

W przypadku, gdy dwa złożone zlecenia mają taki sam kurs, jako pierwsze jest realizowane to zlecenie, które zostało złożone wcześniej.

Składowe zlecenia

1. rodzaj oferty (kupno/sprzedaż)
2. limit ceny (lub informacje o jego braku)
3. termin ważności zlecenia
4. nazwę instrumentu finansowego, który chce kupić/sprzedać
5. wolumen, czyli liczbę papierów wartościowych
6. ewentualne, dodatkowe warunki zlecenia

Najważniejsze cechy Aplikacji

- ▶ innowacyjność
Aplikacja jest pierwszą strumieniową aplikacją przetwarzającą dane giełdowe.
- ▶ efektywność
Dzięki zastosowaniu specjalnych struktur danych do gromadzenia zleceń zminimalizowano złożoność obliczeniową wykonywanych instrukcji.
- ▶ połączenie z dodatkową bazą danych
Aplikacja symuluje system notowań GPW na dwa sposoby:
 - ▶ z wykorzystaniem dedykowanych struktur danych,
 - ▶ z wykorzystaniem zewnętrznej strumieniowej bazy danych.

Funkcjonalności Aplikacji

- ▶ Generator zleceń
- ▶ Symulacja w jednym z dwóch trybów (dedykowanym lub strumieniowym)
- ▶ Analiza regresji

Strumieniowe bazy danych

Aplikacje strumieniowe
Prototypowe DSMS
Problemy otwarte

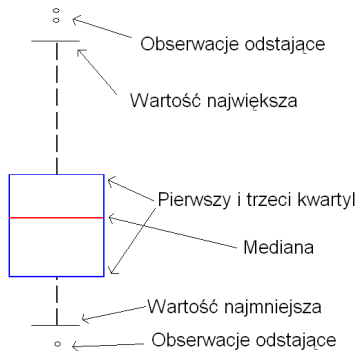
Model systemu giełdowego

WARSET
Zasady funkcjonowania GPW
Aplikacja

Testy porównawcze

Elementy statystyki
Analiza pojedynczych czynników
Analiza regresji wielokrotnej
Testy Wilcoxona
Podsumowanie

Wykres ramkowy



Rysunek: Przykładowy wykres ramkowy

Analiza regresji

Celem analizy regresji było znalezienie liniowej zależności czasu obsługi puli zleceń od ilości zleceń (przy różnych parametryzacjach pozostałych czynników):

$$\text{czas.obslugi} = b_0 + b_1 * \text{liczba.zleceń}$$

Współczynnik b_1 nasi nazwę **współczynnika kierunkowego**,
współczynnik b_0 **wyrazu wolnego**.

Analiza regresji

Jeśli przyjmiemy, że prosta regresji ma postać $y = b_0 + b_1 * x$, wartość $\bar{y}_i = b_0 + b_1 * x_i$ można interpretować jako wartość y przewidywaną na podstawie rozpatrywanej prostej dla wartości zmiennej objaśniającej x równej x_i .

Błąd oszacowania, czyli tzw. **wartość resztowa** lub **rezyduum** wynosi $y_i - \bar{y}_i$.

Analiza regresji

Chcielibyśmy, aby wszystkie wartości rezyduów były jak najmniejsze, co nie jest niestety w praktyce możliwe. W metodzie najmniejszych kwadratów minimalizuje się sumę kwadratów wartości wszystkich odchyleń, czyli rezyduów

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2 = \sum_{i=1}^n (y_i - (b_0 + b_1 * x_i))^2$$

Metodę tę zastosował po raz pierwszy F. Gauss przy analizie danych astronomicznych.

Analiza regresji

Prostą regresji opartą na metodzie najmniejszych kwadratów nazywa się w skrócie prostą regresji MNK lub prostą MNK. Wartości współczynników b_0 i b_1 prostej MNK wynoszą:

$$b_0 = \bar{y} - b_1 \bar{x}$$
$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

gdzie:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Analiza regresji

W celu oceny dobroci dopasowania prostej MNK oblicza się współczynnik SST (*ang. total sum of squares*):

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

Okazuje się, że zmienność opisana przez SST jest sumą dwóch składników, które też mają interpretację zmienności:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \bar{y}_i)^2 + \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 = SSE + SSR$$

Pierwszy ze składników nazywany jest sumą kwadratów błędów (*ang. error sum of squares*), drugi - regresyjną sumą kwadratów (*ang. regression sum of squares*).

Analiza regresji

Stosunek $SSR/SST = 1 - SSE/SST$ zwany **współczynnikiem determinacji (dopasowania)** określa stopień, w jakim zależność liniowa między zmienną objaśnianą a objaśniającą tłumaczy zmienność wykresu rozproszenia.

Analiza regresji

Okazuje się, że wartość współczynnika determinacji (dopasowania) jest ściśle związana z wartością współczynnika korelacji próbkowej R :

$$R^2 = SSR/SST = \frac{\textit{zmiennosc wyjasniana przez model}}{\textit{zmiennosc calkowita}}$$

Wartość R^2 stanowi bardziej adekwatny wskaźnik stopnia zależności liniowej niż sama wartość współczynnika korelacji próbkowej R .

Analiza regresji wielokrotnej

Celem analizy regresji liniowej jest znalezienie funkcji, która opisuje zależność cechy objaśnianej od cech objaśniających. Podobnie jak w przypadku jednej zmiennej objaśniającej, odpowiedniej funkcji poszukuje się minimalizując sumę kwadratów błędów predykcji. W bieżącym podrozdziale estymujemy współczynniki funkcji:

$czas.obslugi =$

$$b_0 + b_1 * liczba.akacji + b_2 * zapelnienie.bufora + b_3 * stosunek.K.S$$

Analiza regresji wielokrotnej

Podczas analizy regresji wielokrotnej sprawdzono każdy z 14 możliwych modeli:

- ▶ czas.obslugi liczba.akcji + zapelnienie.bufora + stosunek.K.S
- ▶ czas.obslugi liczba.akcji + zapelnienie.bufora
- ▶ czas.obslugi liczba.akcji + stosunek.K.S
- ▶ czas.obslugi zapelnienie.bufora + stosunek.K.S
- ▶ czas.obslugi liczba.akcji
- ▶ czas.obslugi zapelnienie.bufora
- ▶ czas.obslugi stosunek.K.S
- ▶ analogiczne modele dla log.czas.obslugi .

Analiza regresji wielokrotnej

Najlepszy współczynnik dopasowania otrzymano dla modelu:

$\log(\text{czas.obslugi}) =$

$b_0 + b_1 * \text{liczba.akacji} + b_2 * \text{zapelnienie.bufora} + b_3 * \text{stosunek.K.S}$

Test Wilcoxona

Średni czas obsługi puli 1000 zleceń dla systemu dedykowanego wynosi ok 400 milisekund, podczas gdy dla strumieniowej bazy danych osiąga wartość 500 milisekund, co sugeruje większą efektywność rozwiązania dedykowanego. Jednakże, dla części obserwacji czas obsługi jest krótszy w przypadku połączenia z zewnętrzną bazą danych.

Test Wilcoxon

W ramach testów porównawczych testowano hipotezę zerową:

$$H_0 : \mu_D = \mu_S$$

przeciwko hipotezie alternatywnej:

$$H_1 : \mu_D < \mu_S$$

gdzie:

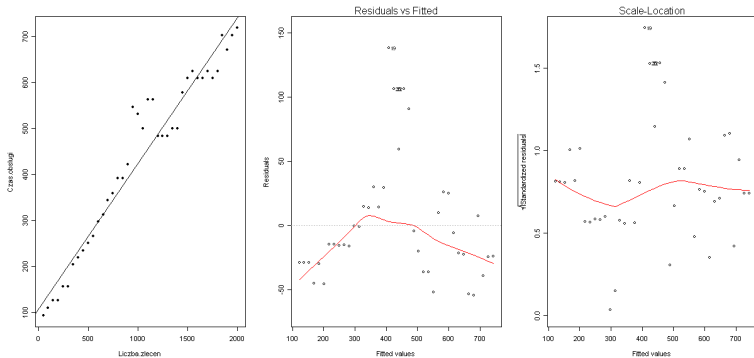
μ_D - średni czas obsługi dla systemu dedykowanego

μ_S - średni czas obsługi dla komercyjnej, strumieniowej bazy danych

Test Wilcoxona

Ze względu na brak założenia o normalności prób testowych, test Wilcoxona ma szerokie zastosowania. Jedynym założeniem jest symetryczność rozkładów (próby pochodzą z tych samych rozkładów przesuniętych względem siebie). Powyższy warunek spełniają wyniki symulacji oparte na systemie dedykowanym i strumieniowej bazie danych.

Pełna pula testowa



Rysunek: Wykres rozproszenia i wykresy rezyduów dla standardowej parametryzacji

Pełna pula testowa

```
> summary(model)

Call:
lm(formula = Czas.obsługi ~ Liczba.zleceń, data = Y)

Residuals:
    Min       1Q   Median       3Q      Max
-54.28 -28.81 -14.81  14.51 138.46

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 105.93846   14.90967    7.105 1.77e-08 ***
Liczba.zleceń  0.31852    0.01267   25.131 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

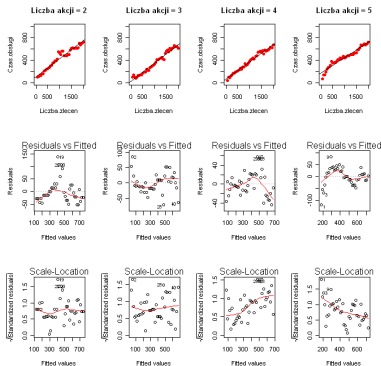
Residual standard error: 46.27 on 38 degrees of freedom
Multiple R-Squared: 0.9432, Adjusted R-squared: 0.9418
F-statistic: 631.5 on 1 and 38 DF, p-value: < 2.2e-16

> anova(model)
Analysis of Variance Table

Response: Czas.obsługi
          Df Sum Sq Mean Sq F value    Pr(>F)
Liczba.zleceń  1 1351917 1351917  631.54 < 2.2e-16 ***
Residuals    38  81345    2141
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

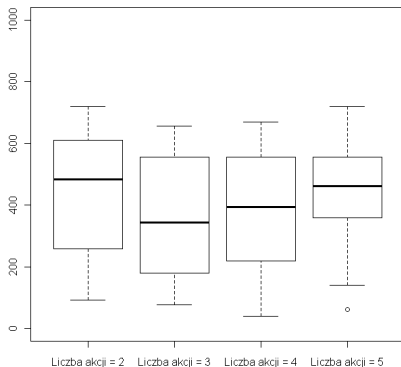
Rysunek: Wydruk z pakietu R z wynikami analizy regresji dla standardowej parametryzacji

Liczba instrumentów



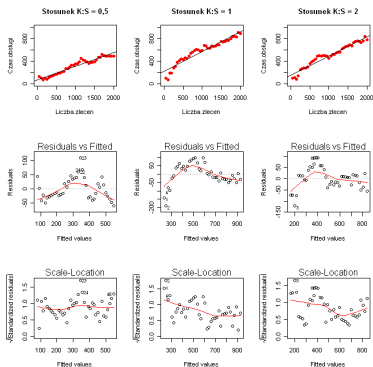
Rysunek: Wykresy rozproszenia i wykresy rezyduów dla zmiennej liczby instrumentów

Liczba instrumentów



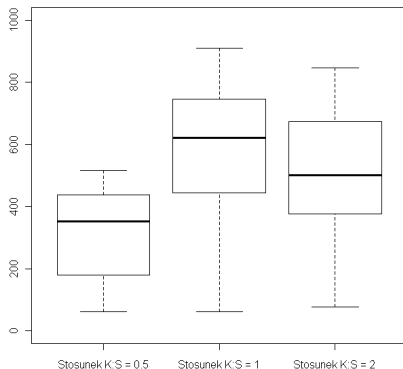
Rysunek: Diagramy ramkowe dla zmiennej liczby instrumentów

Zróznicowanie typów zleceń



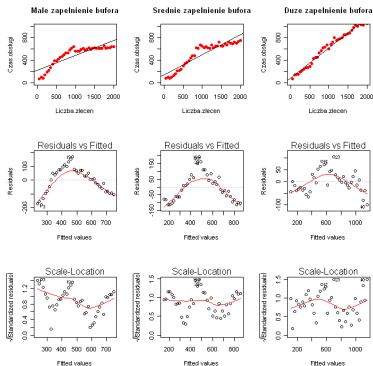
Rysunek: Wykresy rozproszenia i wykresy rezyduów dla zmiennego zróznicowania typów zleceń

Zróżnicowanie typów zleceń



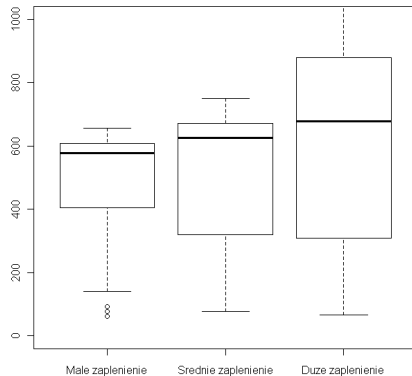
Rysunek: Diagramy ramkowe dla zmiennego zróżnicowania typów zleceń

Zapełnienie bufora zleceń



Rysunek: Wykresy rozproszenia i wykresy rezyduów dla zmiennego zapełnienia bufora

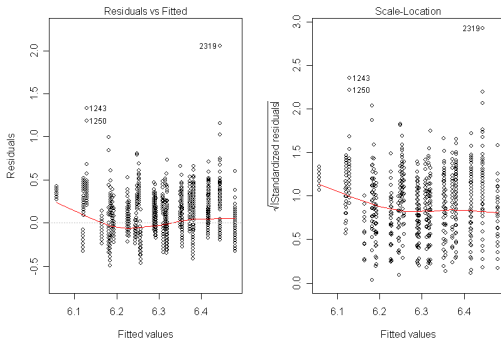
Zapełnienie bufora zleceń



Rysunek: Diagramy ramkowe dla zmiennego zapełnienia bufora

Analiza regresji wielokrotnej

System dedykowany



Rysunek: Wykresy rozproszenia i wykresy rezyduów dla modelu systemu dedykowanego

System dedykowany

```
> summary(model)

Call:
lm(formula = Czas.obsługi ~ ., data = Y)

Residuals:
    Min       1Q   Median       3Q      Max
-0.49371 -0.19553 -0.02979  0.20054  2.05432

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.011614   0.018033   333.37 <2e-16 ***
Liczba.akcji  0.062078   0.003580   17.34 <2e-16 ***
Zapełnienie.bufora 0.064285   0.004902   13.11 <2e-16 ***
Stosunek.K.S -0.071175   0.006418  -11.09 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2401 on 3596 degrees of freedom
Multiple R-Squared:  0.1423, Adjusted R-squared:  0.1414
F-statistic: 198.6 on 3 and 3596 DF, p-value: < 2.2e-16

> anova(model)

Analysis of Variance Table

Response: Czas.obsługi
          Df Sum Sq Mean Sq F value    Pr(>F)
Liczba.akcji      1  17.342   17.342  300.71 < 2.2e-16 ***
Zapełnienie.bufora  1   9.918   9.918  171.99 < 2.2e-16 ***
Stosunek.K.S      1   7.092   7.092  122.98 < 2.2e-16 ***
Residuals        3596 207.374   0.058
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

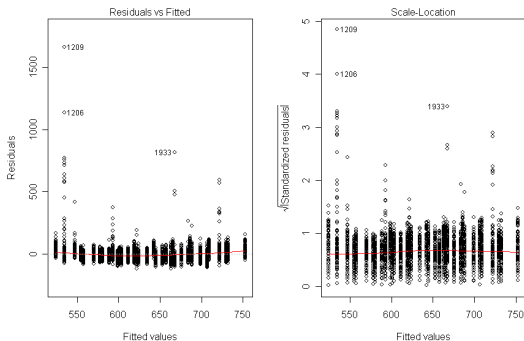
Rysunek: Wydruk z pakietu R z wynikami analizy regresji dla modelu systemu dedykowanego

System dedykowany

	logCzas	liczba akcji	zapelnienie bufora	stosunek K:S
logCzas	1.00	0.27	0.20	-0.17
liczba akcji	0.27	1.00	0.00	0.00
zapelnienie bufora	0.20	0.00	1.00	0.00
stosunek K:S	-0.17	0.00	0.00	1.00

Tabela: Współczynniki korelacji dla modelu systemu dedykowanego

Zewnętrzna bazy danych



Rysunek: Wykresy rozproszenia i wykresy rezyduów dla modelu systemu strumieniowego

Zewnętrzna bazy danych

```
> summary(model)

Call:
lm(formula = Czas.obsługi ~ ., data = Y)

Residuals:
    Min       1Q   Median       3Q      Max
-120.090  -36.352   -7.744   24.485  1663.870

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    403.672     5.321   75.87 <2e-16 ***
Liczba.akcji     22.960     1.056   21.74 <2e-16 ***
Zapelnienie.bufora  64.334     1.446   44.48 <2e-16 ***
Stosunek.K.S     20.536     1.894   10.84 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 70.86 on 3596 degree of freedom
Multiple R-Squared:  0.4167,    Adjusted R-squared:  0.4162
F-statistic: 856.2 on 3 and 3596 DF,    p-value: < 2.2e-16

> anova(model)

Analysis of Variance Table

Response: Czas.obsługi

            Df Sum Sq Mean Sq F value    Pr(>F)
Liczba.akcji      1  2372273  2372273  472.50 < 2.2e-16 ***
Zapelnienie.bufora  1  9933224  9933224  1978.44 < 2.2e-16 ***
Stosunek.K.S      1  590439  590439  117.60 < 2.2e-16 ***
Residuals        3596 18054561    5021
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Rysunek: Wydruk z pakietu R z wynikami analizy regresji dla modelu systemu strumieniowego

Zewnętrzna bazy danych

	logCzas	liczba akcji	zapełnienie bufora	stosunek K:S
logCzas	1.00	0.28	0.57	0.14
liczba akcji	0.28	1.00	0.00	0.00
zapełnienie bufora	0.57	0.00	1.00	0.00
stosunek K:S	0.14	0.00	0.00	1.00

Tabela: Współczynniki korelacji dla modelu systemu strumieniowego

Testy Wilcoxona

Pełna pula testowa

```
> wilcox.test(Y_str, Y_ded, paired = TRUE, alternative = "greater")
```

```
Wilcoxon signed rank test with continuity correction
```

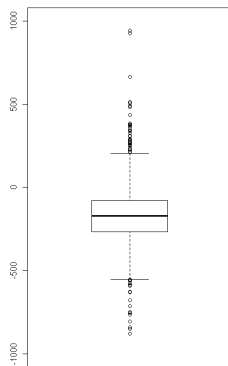
```
data: Y_str and Y_ded
```

```
V = 6079785, p-value < 2.2e-16
```

```
alternative hypothesis: true location shift is greater than 0
```

Rysunek: Wydruk z R z wynikami testu Wilcozona dla pełnej puli danych

Pełna pula testowa



Rysunek: Diagram ramkowy różnicy czasu obsługi dla pełnej puli danych

Liczba instrumentów

```
> wilcox.test(X_str2, X_ded2, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str2 and X_ded2
V = 386937, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

> wilcox.test(X_str3, X_ded3, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str3 and X_ded3
V = 400884, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

> wilcox.test(X_str4, X_ded4, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str4 and X_ded4
V = 336015.5, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

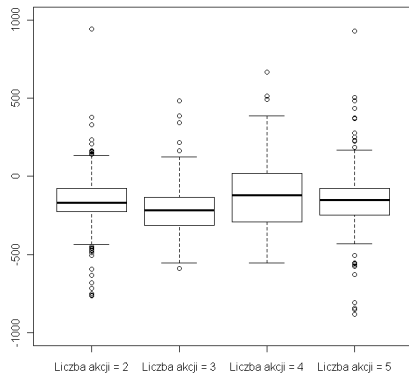
> wilcox.test(X_str5, X_ded5, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str5 and X_ded5
V = 384147, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0
```

Rysunek: Wydruk z R z wynikami testu Wilcozona dla zmiennej liczby akcji

Liczba instrumentów



Rysunek: Diagram ramkowy różnicy czasu obsługi dla zmiennej liczby akcji

Zróznicowanie typów zleceń

```
> wilcox.test(X_str1, X_ded1, paired = TRUE, alternative = "greater")

      Wilcoxon signed rank test with continuity correction

data:  X_str1 and X_ded1
V = 664289, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

> wilcox.test(X_str2, X_ded2, paired = TRUE, alternative = "greater")

      Wilcoxon signed rank test with continuity correction

data:  X_str2 and X_ded2
V = 647755.5, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

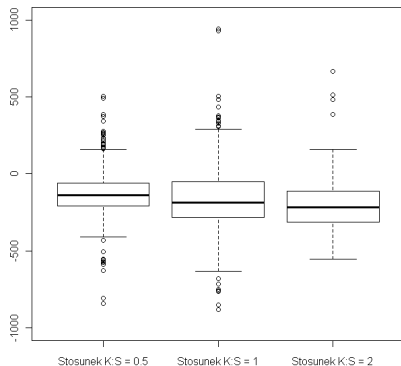
> wilcox.test(X_str3, X_ded3, paired = TRUE, alternative = "greater")

      Wilcoxon signed rank test with continuity correction

data:  X_str3 and X_ded3
V = 709726, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0
```

Rysunek: Wydruk z R z wynikami testu Wilcozona dla zmiennego zróznicowania typów zleceń

Zróżnicowanie typów zleceń



Rysunek: Diagram ramkowy różnicy czasu obsługi dla zmiennego zróżnicowania typów zleceń

Zapełnienie bufora zleceń

```
> wilcox.test(X_str1, X_ded1, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str1 and X_ded1
V = 643133.5, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

> wilcox.test(X_str2, X_ded2, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str2 and X_ded2
V = 694917.5, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0

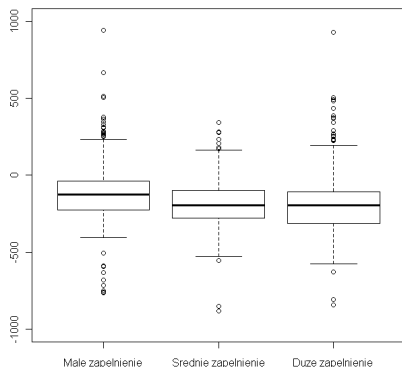
> wilcox.test(X_str3, X_ded3, paired = TRUE, alternative = "greater")

Wilcoxon signed rank test with continuity correction

data: X_str3 and X_ded3
V = 682867, p-value < 2.2e-16
alternative hypothesis: true location shift is greater than 0
```

Rysunek: Wydruk z R z wynikami testu Wilcozona dla zmiennego zapełnienia bufora

Zapełnienie bufora zleceń



Rysunek: Diagram ramkowy różnicy czasu obsługi dla zmiennego zapełnienia bufora

Koniec

Dziękuję za uwagę.