

# Zadanie zaliczeniowe OVB

Aleksander Gutkowski

17 września 2021

## Spis treści

<b>1</b>	<b>Zadanie</b>	<b>1</b>
1.1	Wstęp - opis zadania . . . . .	1
1.2	Opis danych . . . . .	1
1.3	Wstępny wygląd danych oraz pomysł na zadanie . . . . .	2
1.4	Pierwszy problem . . . . .	4
1.5	Sumowanie danych, maxima oraz minima . . . . .	4
1.6	Średnia, średnia trymowana i kolejne problemy . . . . .	6
1.7	Nowy problem - logiczny . . . . .	10
1.8	Punkt stabilności . . . . .	11
1.9	Wykresy . . . . .	11
1.10	Dygresja . . . . .	13
1.11	Kolejne wykresy . . . . .	14
1.12	Kolejna dygresja . . . . .	16
1.13	Odchylenie standardowe - po co? . . . . .	17
1.14	Naj <sup>2</sup> . . . . .	18
1.15	Kolejny estymator - współczynnik zmienności . . . . .	19
1.16	Procentowy udział firm . . . . .	20
<b>2</b>	<b>Podsumowanie</b>	<b>20</b>

## 1 Zadanie

### 1.1 Wstęp - opis zadania

Jako zadanie zaliczeniowe, oraz jako zadanie, by przećwiczyć zdobyte umiejętności podczas nauki na studiach, do "obróbki" dostałem dane dotyczące zdobytych punktów przez osobę w firmie. Na podstawie danych postaram się zrozumieć jakie zależności tu występują oraz podsumować zdobytą wiedzę z tych liczb.

### 1.2 Opis danych

Do przejżenia dostałem dane od maja 2018 roku do sierpnia 2021 roku. Firm, z których zdobyto punkty za umowy jest 15 np. Allianz, Generali, Prudential. Dodatkowo rozwinę skróty - nazwy firm:

1. AMP - MetLife
2. ALH - Hestia życie
3. PRU - Prudential
4. ALL - Allianz życie
5. SID - Signal Induna życie
6. WAR - Warta
7. NNN - Nationale Nederlanden
8. CMM - Compensa majątek
9. GEN - Generali życie
10. GEM - Generali majątek
11. LEG - Esaliens
12. ALM - Allianz majątek
13. HEI - Hestia majątek
14. SIM - Signal majątek
15. DAS - ochrona prawna

### **1.3 Wstępny wygląd danych oraz pomysł na zadanie**

Do przedstawienia wszystkich danych, jako pierwsza myśl przyszło mi przedstawienie wektorowe - jako miesiąc jeden wektor np  $[0, 0, 12, 0, 24 \dots]$ , jednak pojawia się problem, bo trzeba by było stworzyć tyle wektorów ile jest miesięcy - 40. Gdyby odwrócić sytuację to było by trochę lepiej, bo "tylko" 15 wektorów - ilość firm. Dlatego moim pomysłem na przedstawienie tego w sprawny sposób jest macierz, a teraz przedstawię jak wygląda cała struktura.

*# ?matrix - do sprawdzenia jak zapisać macierz oraz jakie można zrobić na niej operacje*

```
dane = matrix(c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 20, 0, 0, 0, 0,
0, 82, 0, 84, 0, 0, 0, 0, 14, 0, 0, 0, 0, 2, 0,
0, 25, 0, 145, 44, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 163, 0, 33, 0, 0, 0, 12, 0, 0, 0, 0, 0, 15, 0,
0, 302, 0, 0, 0, 52, 0, 20, 0, 2, 0, 0, 0, 0, 0,
0, 40, 0, 90, 58, 0, 0, 7, 14, 0, 0, 0, 0, 22, 0,
0, 189, 0, 147, 123, 15, 259, 19, 21, 21, 0, 0, 0, 22, 0,
0, 204, 0, 118, 29, 34, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 203, 0, 0, 110, 5, 0, 0, 16, 0, 20, 0, 0, 0, 0,
107, 0, 178, 296, 60, 0, 0, 8, 17, 0, 0, 0, 0, 0, 0,
0, 0, 44, 0, 42, 0, 0, 13, 35, 0, 0, 0, 0, 0, 0,
54, 95, 0, 247, 24, 8, 0, 14, 44, 0, 40, 0, 7, 0, 0,
0, 50, 307, 94, 75, 61, 0, 7, 0, 2, 0, 0, 9, 0, 0,
0, 0, 126, 48, 44, 18, 0, 4, 14, 2, 40, 23, 0, 4, 0,
0, 121, 203, 40, 44, 36, 0, 61, 0, 4, 20, 0, 0, 0, 0,
144, 33, 427, 42, 160, 74, 0, 28, 200, 8, 40, 0, 0, 3, 7,
167, 0, 219, 32, 114, 25, 0, 95, 32, 12, 20, 0, 0, 0, 0,
0, 60, 251, 222, 98, 12, 0, 70, 83, 30, 0, 0, 47, 0, 0,
0, 0, 82, 0, 0, 22, 0, 47, 16, 36, 0, 0, 6, 0, 0,
0, 67, 109, 0, 20, 120, 0, 14, 51, 34, 0, 24, 17, 0, 0,
22, 132, 417, 203, 78, 17, 0, 26, 68, 57, 0, 0, 0, 0, 0,
79, 188, 965, 529, 82, 140, 0, 57, 45, 68, 60, 0, 26, 0, 0,
670, 492, 0, 113, 224, 27, 0, 87, 19, 30, 80, 3, 27, 0, 0,
144, 0, 0, 0, 78, 0, 0, 0, 60, 26, 0, 0, 0, 0, 0,
662, 109, 0, 34, 17, 17, 0, 29, 41, 56, 0, 0, 0, 0, 0,
323, 127, 0, 0, 0, 32, 0, 36, 0, 43, 0, 20, 9, 0, 0,
718, 43, 109, 13, 12, 79, 0, 62, 0, 92, 0, 0, 0, 0, 0,
523, 0, 137, 5, 0, 91, 0, 109, 69, 11, 40, 0, 0, 0, 0,
130, 125, 0, 60, 0, 67, 0, 170, 275, 55, 20, 0, 0, 0, 0,
343, 0, 0, 0, 104, 10, 0, 37, 18, 40, 60, 0, 0, 0, 0,
207, 105, 0, 63, 57, 77, 0, 73, 52, 33, 20, 12, 17, 0, 0,
52, 165, 0, 0, 0, 105, 0, 60, 18, 77, 20, 59, 29, 0, 0,
267, 255, 103, 0, 18, 45, 0, 20, 19, 43, 11, 26, 0, 0, 0,
416, 441, 90, 46, 18, 102, 0, 67, 0, 27, 31, 95, 5, 0, 0,
49, 369, 0, 0, 0, 104, 1172, 92, 66, 33, 23, 28, 19, 0, 0,
366, 284, 100, 133, 0, 8, 0, 0, 0, 31, 63, 0, 12, 0, 0,
305, 153, 365, 28, 35, 58, 0, 13, 42, 176, 11, 86, 4, 0, 0,
278, 0, 0, 0, 0, 0, 0, 36, 14, 72, 11, 77, 23, 0, 0,
109, 0, 0, 0, 0, 0, 0, 8, 0, 31, 0, 0, 0, 0, 0), nrow = 40, ncol = 15, byrow = TRUE)
```

## 1.4 Pierwszy problem

Jak widać jest bardzo dużo danych. Teraz pytanie co z nimi zrobić? Na dobry początek sprawdzę, ile razem punktów zdobyła dana firma.

	AMP	ALH	PRU	ALL	SID
Suma	6135	4622	4232	2865	1768

Tabela 1: Suma punktów za dane firmy (5)

	WAR	NNN	CMM	GEN	GEM
Suma	1461	1431	1401	1363	1152

Tabela 2: Suma punktów za dane firmy (5)

	LEG	ALM	HEI	SIM	DAS
Suma	650	459	257	68	7

Tabela 3: Suma punktów za dane firmy (5)

Jak widać różni się to od tego co jest podane w Excelu. Dysproporcja polega na tym że, dane w programie R są naturalne, a w Excelu - wymierne (części dziesiętne, setne ...). Sprawdźmy, jaka jest największa różnica między podanym wynikiem w Excelu a R'em:

```
## [1] 3
```

jest to bardzo niska liczba. Wiedząc, że ten wynik dot. 3 firm: AMP, ALL oraz SID, błąd ten przedstawię w tabeli:

	AMP	ALL	SID
Różnica [%]	0.05	0.10	0.17

Tabela 4: Tabela różnicy sum firm przedstawiona procentowo

Jak można zobaczyć, jest to błąd na poziomie  $> 1\%$ , dlatego uważam, że dane na których będę operować w programie R, będą odpowiednie.

## 1.5 Sumowanie danych, maxima oraz minima

Skoro mamy wyliczone już, ile mamy w sumie punktów za każdą firmę, sprawdźmy która z nich ma najwięcej punktów.

```
## [1] 6135
```

Jak można odczytać z Tabeli 1 jest to firma AMP - MetLife. Jest to spowodowane prawdopodobnie dobrym produktem - Gwarancja Jutra. Teraz sprawdzę najmniej punktów.

## [1] 7

Punkty te pochodzą z firmy DAS (Tabela 3). Spowodowane jest to zapewne tym, że nie dużo osób potrzebuje ochrony prawnej + firma odeszła z Polski.

Następnie sprawdźmy ile też jest wszystkich punktów w sumie.

## [1] 27871

Teraz stworzę tabelę dla miesięcy i otrzymanych w nich punktów.

	05.2018	06.2018	07.2018	08.2018	09.2018
Suma	6	20	182	214	223

Tabela 5: Suma punktów za dane miesiące (5)

	10.2018	11.2018	12.2018	01.2019	02.2019
Suma	376	231	816	385	354

Tabela 6: Suma punktów za dane miesiące (5)

	03.2019	04.2019	05.2019	06.2019	07.2019
Suma	666	134	533	605	323

Tabela 7: Suma punktów za dane miesiące (5)

	08.2019	09.2019	10.2019	11.2019	12.2019
Suma	529	1166	716	873	209

Tabela 8: Suma punktów za dane miesiące (5)

	01.2020	02.2020	03.2020	04.2020	05.2020
Suma	456	1020	2239	1772	308

Tabela 9: Suma punktów za dane miesiące (5)

	06.2020	07.2020	08.2020	09.2020	10.2020
Suma	965	590	1128	985	902

Tabela 10: Suma punktów za dane miesiące (5)

	11.2020	12.2020	01.2021	02.2021	03.2021
Suma	612	716	585	807	1338

Tabela 11: Suma punktów za dane miesiące (5)

	04.2021	05.2021	06.2021	07.2021	08.2021
Suma	1955	997	1276	511	148

Tabela 12: Suma punktów za dane miesiące (5)

Skoro wiemy, które z firm dały najwięcej oraz najmniej punktów, sprawdźmy które miesiące dały najwięcej oraz najmniej punktów.

```
## [1] 2239
```

Z tabeli 9 można odczytać, iż jest to marzec 2020. Przypuszczam, że ma to związek z pandemią - był to początek epidemii na świecie. Teraz najniższy wynik:

```
## [1] 6
```

Jest to maj 2018. Wynik ten może być spowodowany początkiem pracy oraz zakończeniem miesiąca (np. zrobienie pierwszej umowy, małej ostatniego dnia miesiąca). Dodatkowo sprawdźmy najwyższy i najniższy pojedynczy wynik.

```
## [1] 1172
## [1] "^ najwyższy wynik - NNN w kwietniu 2021 roku"
## [1] 2
## [1] "^ najniższy wynik i pojawia się ono w 4 miejscach"
```

Tak wysoki wynik w NNN - Nationale Nederlanden, spowodowany był dosyć dużą grupówką.

## 1.6 Średnia, średnia trymowana i kolejne problemy

Sprawdźmy kolejną ważną rzecz - średnia. Zaczniemy od sprawdzenia średniej ilości zdobytych punktów przez podaną firmę:

	AMP	ALH	PRU	ALL	SID
Średnia	153	116	106	72	44

Tabela 13: Średnia punktów za dane firmy (5)

	WAR	NNN	CMM	GEN	GEM
Średnia	37	36	35	34	29

Tabela 14: Średnia punktów za dane firmy (5)

	LEG	ALM	HEI	SIM	DAS
Średnia	16	11	6	2	0

Tabela 15: Średnia punktów za dane firmy (5)

Jak widać, firma DAS ma średnią równą 0. Czemu? Jest to spowodowane tym, że program zebrał wszystkie wyniki (w tym 0 - kiedy nie było żadnych punktów za umowę) i w konsekwencji wynik jest zniekształcony, dlatego trzeba sprawdzić, jak wygląda prawdziwa średnia (zaokrąglona to jedność):

	AMP	ALH	PRU	ALL	SID
Średnia	267	165	235	115	68

Tabela 16: Średnia punktów za dane firmy (5)

	WAR	NNN	CMM	GEN	GEM
Średnia	50	716	44	50	40

Tabela 17: Średnia punktów za dane firmy (5)

	LEG	ALM	HEI	SIM	DAS
Średnia	32	38	17	11	7

Tabela 18: Średnia punktów za dane firmy (5)

a teraz średnia, ale ze względu na miesiąc:

	05.2018	06.2018	07.2018	08.2018	09.2018
Średnia	6	20	46	71	56

Tabela 19: Średnia punktów za dane miesiące (5)

	10.2018	11.2018	12.2018	01.2019	02.2019
Średnia	94	38	91	96	71

Tabela 20: Średnia punktów za dane miesiące (5)

	03.2019	04.2019	05.2019	06.2019	07.2019
Średnia	111	34	59	76	32

Tabela 21: Średnia punktów za dane miesiące (5)

	08.2019	09.2019	10.2019	11.2019	12.2019
Średnia	66	97	80	97	35

Tabela 22: Średnia punktów za dane miesiące (5)

	01.2020	02.2020	03.2020	04.2020	05.2020
Średnia	51	113	204	161	77

Tabela 23: Średnia punktów za dane miesiące (5)

	06.2020	07.2020	08.2020	09.2020	10.2020
Średnia	121	84	141	123	113

Tabela 24: Średnia punktów za dane miesiące (5)

	11.2020	12.2020	01.2021	02.2021	03.2021
Średnia	87	65	65	81	122

Tabela 25: Średnia punktów za dane miesiące (5)

	04.2021	05.2021	06.2021	07.2021	08.2021
Średnia	196	125	106	73	49

Tabela 26: Średnia punktów za dane miesiące (5)

Teraz posiadając tak wiele danych warto znaleźć dla jakiej firmy, jest najwyższa średnia:

```
## [1] 716
```

Średnia ta należy do NNN, jednak jest to wynik niedokładny, do którego wrócę później. Następne w kolejności jest najniższa średnia:

```
## [1] 7
```

i należy ona do firmy DAS. Było to do przewidzenia ze względu na to, że ilość miesięcy, w których zdobyto punkty wynosi 1.

Dla miesięcy wygląda to odpowiednio:

```
## [1] 204
```

dla marca 2020 roku i

```
## [1] 6
```

dla maja 2018 roku.

Powracając do wcześniej powstałego problemu. Najwyższa średnia dla NNN jest zakłamaną. Jak to możliwe, skoro było to obliczane przez komputer. Otóż dzięki wcześniejszym obliczeniom widać, że NNN uzyskało jednorazowo najlepszy wynik, a przez to że NNN miało bardzo małą ilość miesięcznych zdobytych punktów, nie jest to poprawny do końca wynik. Jak więc uzyskać najlepszy wynik? Do tego przyda się średnia trymowana, działająca pod inną też, jako średnia ucinana. Jak się to liczy? Bardzo prosto - liczymy tak jak średnią jednak nie bierzemy pod uwagę najniższej oraz najwyższej wartości. Po co się to robi? Średnia ucinana jest mało wrażliwa na wartości odstające - bo usunęliśmy je i zastąpiliśmy liczbami bliższymi mediany.



Aby policzyć średnią trymowaną dla poszczególnych firm, trzeba spełnić warunek. Ilość miesięcy, w których zdobyto punkty musi być większa niż 2. Sprawdźmy w takim razie, które z firm spełniają warunek.

```
## [1] 23
## [1] "^ AMP"
## [1] 28
## [1] "^ ALH"
## [1] 18
## [1] "^ PRU"
## [1] 25
## [1] "^ ALL"
## [1] 26
## [1] "^ SID"
## [1] 29
## [1] "^ WAR"
## [1] 2
## [1] "^ NNN"
## [1] 32
## [1] "^ CMM"
## [1] 27
## [1] "^ GEN"
## [1] 29
## [1] "^ GEM"
## [1] 20
## [1] "^ LEG"
## [1] 12
## [1] "^ ALM"
## [1] 15
## [1] "^ HEI"
## [1] 6
## [1] "^ SIM"
## [1] 1
## [1] "^ DAS"
```

Jak widać do obliczenia nie kwalifikują się tylko 2 firmy: NNN i DAS. Teraz policzę średnią ucinaną, ale najpierw kod:

```
srednia_ucinana1 = matrix(c(dane2[,1]))
srednia_ucinana1[srednia_ucinana1 == 0] = NA
srednia_ucinana2=(sort(srednia_ucinana1))
srednia_ucinana3 = srednia_ucinana2[2:22]
round(mean(srednia_ucinana3))

## [1] 257
```

Jak obliczyłem to? Kod ma kilka kroków, na których wpadnięcie trochę mi zeszło czasu. Pierwszy z nich to wyciągnięcie pierwszego pionowego wektora - punkty z firmy AMP przez cały okres 40 miesięcy. Następnie tam gdzie w danym miesiącu nie zdobyto żadnych punktów czyli

0 zamieniłem na wartość NA - nie dotyczy (not applicable). Następnie sortuje wszystkie wyniki od malejącego do rosnącego. Dzięki wcześniejszej zamianie wszystkie wyniki, które mają NA nie są brane pod uwagę - nie psuje to ucinana skrajnych liczb. Następnie wiedząc z wcześniejszych obliczeń ile jest miesięcy, w których zdobyto punkty, "wycinam" wszystkie, oprócz skrajnych. Z wyciętego skrawka liczę średnią, a na końcu zaokrąglam to do jedności. Czy coś to da? Przekonamy się po obliczeniu wszystkich możliwych średnich wyników punktowych w zależności od firmy. Na ten moment już wiadomo, że żadne z wcześniej wymienionych firm - NNN oraz DAS, które posiadały odpowiednio najwyższą oraz najniższą średnią, nie będą brane pod uwagę - brak spełnienia jednego warunku.

	AMP	ALH	PRU	ALL	SID
Średnia ucinana	257	47	201	101	64

Tabela 27: Średnia trymowana punktów za dane firmy (5)

	WAR	CMM	GEN	GEM
Średnia ucinana	46	41	43	34

Tabela 28: Średnia trymowana punktów za dane firmy (4)

	LEG	ALM	HEI	SIM
Średnia ucinana	31	36	16	11

Tabela 29: Średnia trymowana punktów za dane firmy (4)

teraz dzięki tabelom zobaczymy, które z firm "dały" najwięcej oraz najmniej punktów. Najwięcej punktów:

```
## [1] 257
```

jest to firma AMP, możemy to odczytać z tabeli numer 27, a najmniej:

```
## [1] 11
```

SIM, której wynik jest w tabeli 29. Wniosek nasuwa się sam. Mimo, iż najwyższa średnia była w NNN to było to spowodowane tylko jednym miesiącem. Ten sam wniosek tyczy się DAS. Średnia trymowana w pewien sposób podpowiada, gdzie, w tym przypadku, są najbardziej stałe zdobywanie punktów, co za tym idzie, która firma jest najlepsza, która najlepiej się sprzedaje. Teraz jako przejście do kolejnego segmentu przedstawię wyniki danych firm, oraz okresów jako wykresy.

## 1.7 Nowy problem - logiczny

Ilość wykresów może być bardzo duża, aż za dużo, a w tej pracy nie chodzi o ilość ale o jakość. Dlatego by nie robić 40 wykresów z każdego miesiąca, (albo 15 firm) zrobię wykres pierwszego, ostatniego miesiąca oraz od momentu stabilizacji co pół roku. Czym w takim razie będzie okres stabilizacji? Według mnie jest to moment, podczas którego wyniki z poprzedniego miesiąca są

wyższe niż z aktualnego. Nie jest to najlepszy sposób przedstawienia stabilizacji i znalezienia go, bo np mógł się wydarzyć jakiś wypadek, który zakłócił rośnięcie miesięczne punktów. Dlatego gdy znajdziemy podejrzany moment stabilizacji sprawdzimy czy w kolejnym imiesiącu zwiększy się liczba punktów o 50 %. Sprawdźmy więc, za pomocą tabel 5 - 12.

## 1.8 Punkt stabilności

Z tabel wynika, że pierwszym momentem, gdy liczba punktów za miesiąc była niższa niż w poprzednim była w listopadzie 2018 roku. Teraz zobaczmy czy spełnia ww. warunek:

```
## [1] 71.69118
```

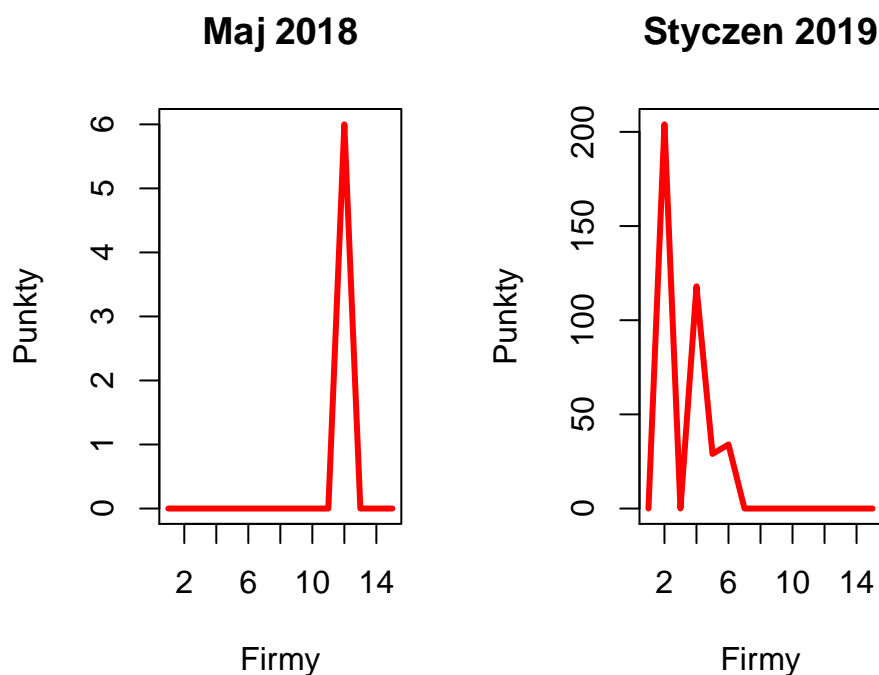
jak widać nie bo wynik wyszedł na poziomie  $> 70\%$ . Kolejny moment to styczeń 2019 rok.

```
## [1] -8.757062
```

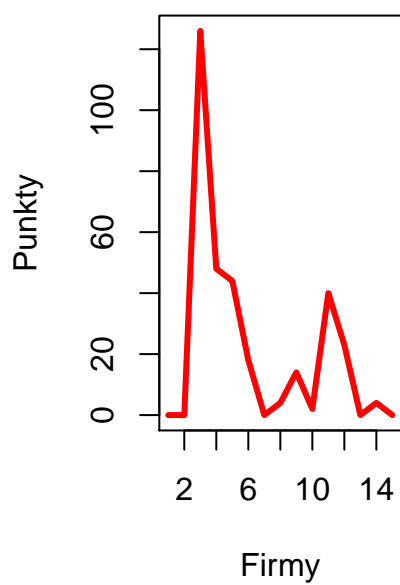
jest to moment stabilności. Jak można zobaczyć kolejny miesiąc jest z dużym zyskiem punktowym. Gdyby było dużo więcej danych - więcej miesięcy można by było zrobić punkt stabilności na długość 3 miesięcy, problem w tym, że wyniki cały czas się zwiększają, co oznacza, że osoba cały czas się rozwija. Mając już punkt stabilności - styczeń 2019 wykonamy parę wykresów, by zobaczyć, jak osoba rozwijała się co pół roku.

## 1.9 Wykresy

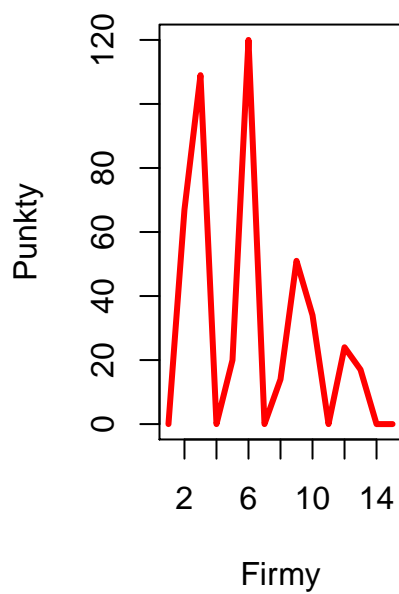
Tak jak wcześniej pisałem, wykresów będzie sporo, jednak jeśli chodzi o miesiące skupimy się na paru - pierwszy, ostatni, punkt stabilności i co pół roku od tego momentu



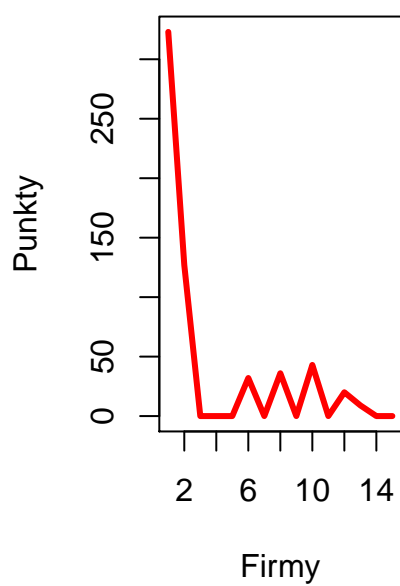
**Lipiec 2019**



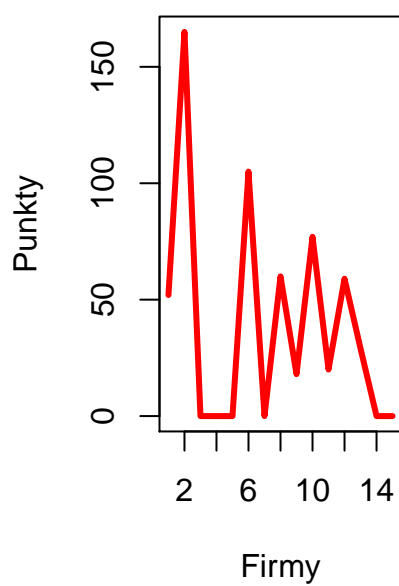
**Styczen 2020**

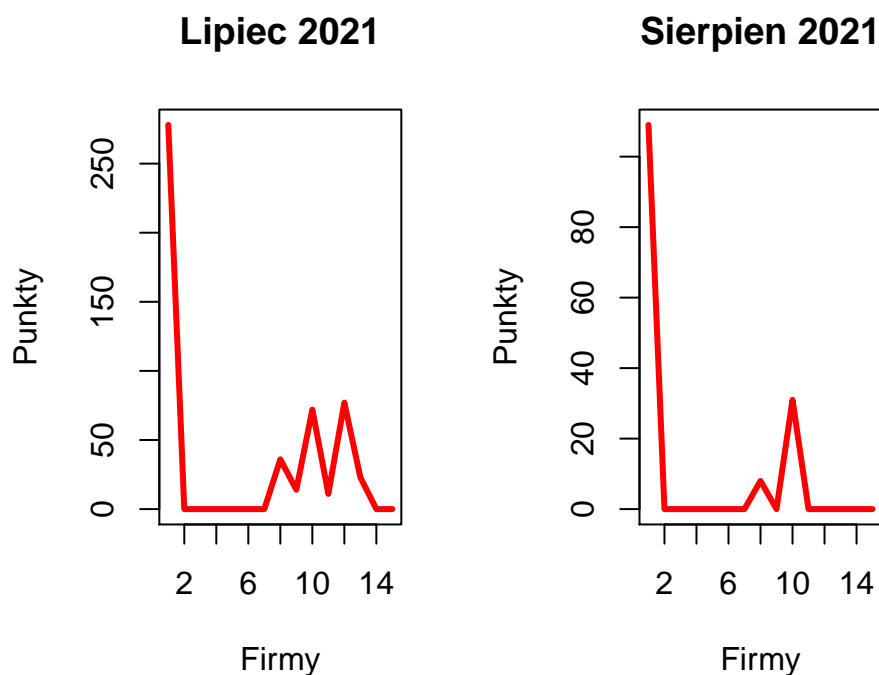


**Lipiec 2020**



**Styczen 2021**





Co można wywnioskować z wykresów? W sumie wszystko to co wcześniej było powiedziane, jednak wizualizacja pomaga zrozumieć pewne rzeczy tj.

- Wykres w większości ma wierzchołek bliżej lewej strony - firmy typu AMP ALH oraz PRU zdobywają więcej punktów niż SIM lub DAS;
- Początkowe okresy mają dużo mniej punktów zdobytych niż późniejsze - rozwój;
- Dzięki wykresom można też zobaczyć, że na prawie wszystkich wykresach jest jeden wierzchołek - 1 firma dominowała punktowo. W styczniu 2020 było inaczej - są 2 wierzchołki oraz żadna z dominujących firm - AMP, ALH, PRU, nie zdobyła tych punktów. Prowadzi nas to do pytania - co to za firma oraz co było powodem takiego wzrostu?

## 1.10 Dygresja

Jak pisałem wcześniej przyjrzymy się styczniowi 2020 roku, aby zbadać pewne "zaburzenie" w wykresie.

```
## [1] 0 67 109 0 20 120 0 14 51 34 0 24 17 0 0
```

Mamy wypis wszystkich wyników, które zdobyto w danym terminie. Teraz, aby lepiej to wyglądało zrobimy z tego tabelkę, ale najpierw sprawdzimy dodatkowo ile punktów było w tym miesiącu razem:

```
## [1] 456
```

Dzięki temu oraz sprawdzeniu tabel 8 oraz 9 widać, że nie był to najłpszy wynik w ciągu ostatnich 3 miesięcy wstecz i w przód. Możliwe, że przez ogólny wynik wykres jest zaburzony. Aby sprawdzić która firma dokonała tego zrobimy wcześniej wspomnianą tabelę.

	AMP	ALH	PRU	ALL	SID
Punkty	0	67	109	0	20

Tabela 30: Punkty za styczeń 2020 dla firm (5)

	WAR	NNN	CMM	GEN	GEM
Punkty	120	0	14	51	34

Tabela 31: Punkty za styczeń 2020 dla firm (5)

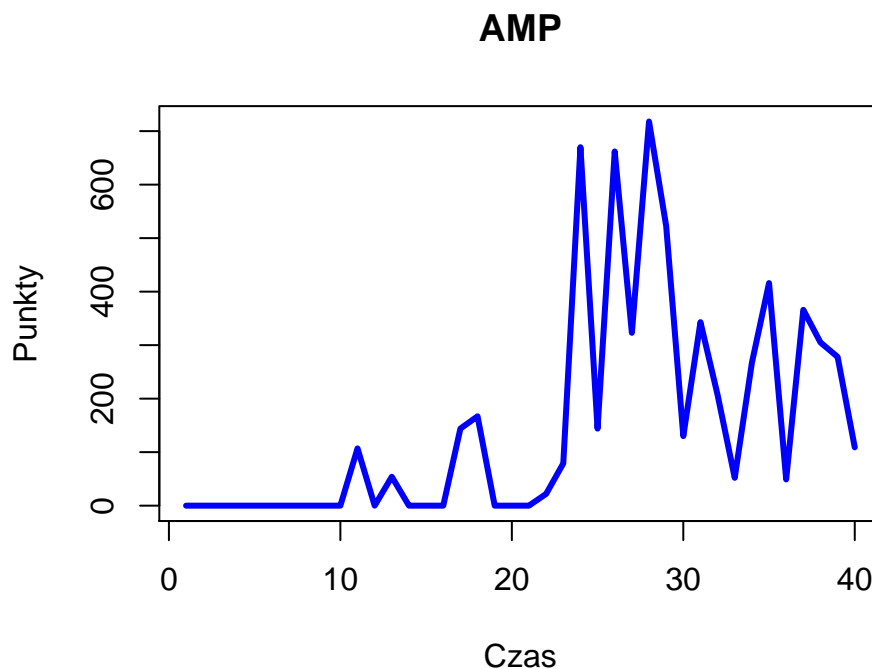
	LEG	ALM	HEI	SIM	DAS
Punkty	0	24	17	0	0

Tabela 32: Punkty za styczeń 2020 dla firm (5)

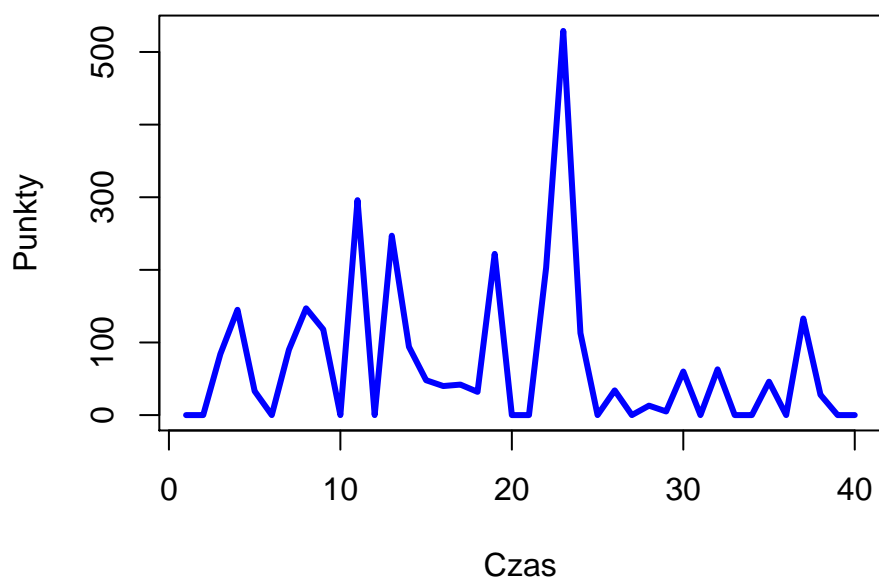
Firmą odpowiadającą za tę anomalię była WAR. Teraz warto zastanowić się czym było to spowodowane super produktem firmy czy słabym obrotem reszty firm.

### 1.11 Kolejne wykresy

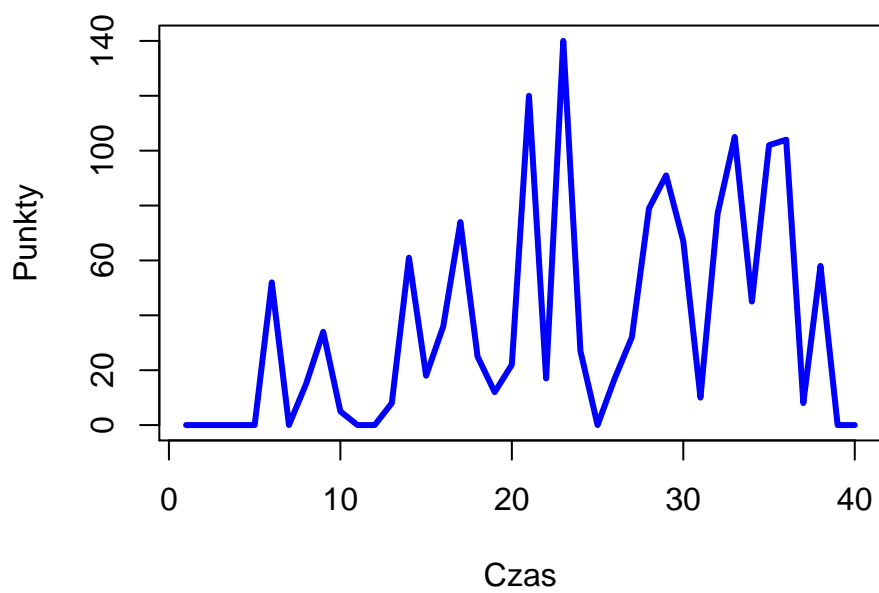
Dzięki dygresji wiemy, że warto przyglądać się takim zjawiskom i analizować je, za pomocą wcześniejszych tabel oraz wyników. Teraz zrobię wykresy, jednak dla firm, względem czasu. Aby nie robić wielu wykresów (15), stwórzę dla 4 wybranych firm - AMP, ALL, WAR oraz CMM. Wybrałem je ze względu na wysoką ilość punktów, ale nie chciałem wybrać tylko 4 pierwszych oraz mają przynajmniej ponad połowę okresów ze zdobytymi punktami (więcej miesięcy w których były zdobyte punkty, niż te bez punktów.)

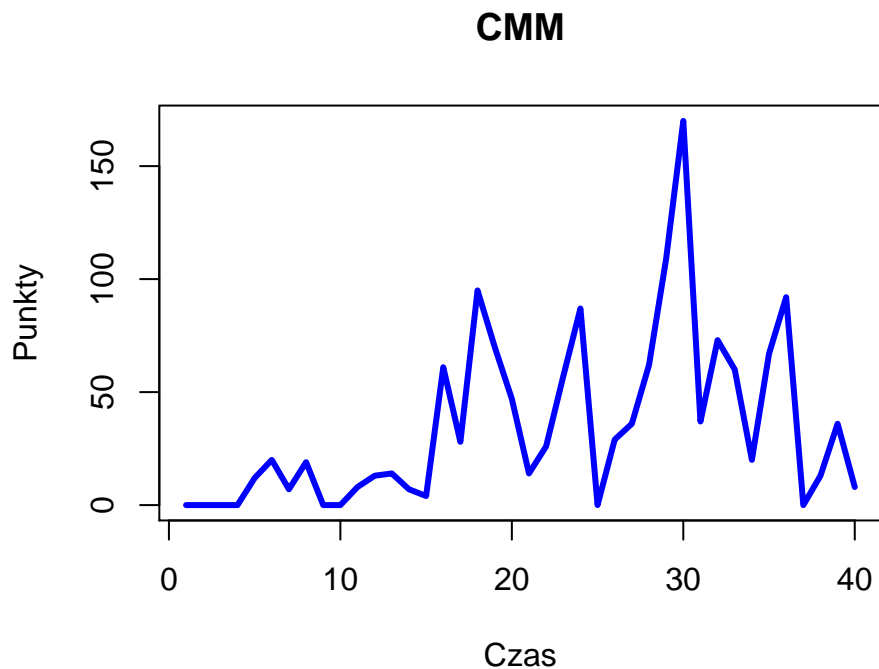


### ALL



### WAR





Jak widać, wszystkie wykresy mają w bardzo podobnych miejscach wierzchołki, oprócz jednego - CMM.

## 1.12 Kolejna dygresja

3 firmy - AMP, ALL, WAR mają wierzchołki w podobnych miejscach - początek pandemii. Jednak firma CMM ma wierzchołek w innym miejscu. Za pomocą wykresu widać, że jest to 30 miesiąc, jednak by mieć pewność sprawdzimy to za pomocą tabeli:

	06.2020	07.2020	08.2020	09.2020	10.2020
Punkty	29	36	62	109	170

Tabela 33: Punkty w firmie CMM w danych miesiącach (5)

	11.2020	12.2020	01.2021	02.2021	03.2021
Punkty	37	73	60	20	67

Tabela 34: Punkty w firmie CMM w danych miesiącach (5)

Jak widać 30 miesiąc - październik 2020, jest najlepszym miesiącem pod względem CMM punktowo. Spowodowane jest to tym, że CMM nie "dawało" dużo punktów, jednak w tym miesiącu było inaczej. Dodatkowo dla 3 pozostałych firm widać z wykresu, że w tym miesiącu miały one niższe wyniki. Warto porównać ile w tym okresie zdobyły pozostałe firmy, by wiedzieć czy te 170 punktów to wiele na tle pozostałych. Teraz przedstawię to w formie tabeli.

	AMP	ALL	WAR	CMM
Punkty	130	60	67	170

Tabela 35: Porównanie punktów z października 2020 dla 4 firm



Jak widać firmy nie odbiegają od siebie w znacznym stopniu. Można to zobaczyć za pomocą procentowej różnicy

	AMP	ALL	WAR	CMM
Różnica [%]	23.53	64.71	60.59	0.00

Tabela 36: Porównanie punktów z października 2020 dla 4 firm

Przedstawienie procentowe pokazuje, że jest to różnica około 60% dla dwóch firm oraz 24% dla AMP. Uważam że nie jest to dużo.

### 1.13 Odchylenie standardowe - po co?

Może się wydawać, że nowe pojęcia które wprowadzam nie mają zbytnio sensu. Jednak odchylenie standardowe jest bardzo użyteczne w statystyce. Dzięki temu też, obok średniej trymowanej, pomaga znaleźć jak najbardziej stabilne dane. Według Wikipedii : "odchylenie standardowe mówi, jak szeroko wartości jakiejś wielkości (na przykład wieku, inflacji, kursu walutowego) są rozrzucone wokół jej średniej. Im mniejsza wartość odchylenia tym obserwacje są bardziej skupione wokół średniej." Sprawdźmy więc jak to ma się do naszych zdobytych punktów:

	AMP	ALH	PRU	ALL	SID
Odchylenie standardowe	207	127	186	107	52

Tabela 37: Odchylenie standardowe punktów za dane firmy (5)

	WAR	NNN	CMM	GEN	GEM
Odchylenie standardowe	40	189	38	54	35

Tabela 38: Odchylenie standardowe punktów za dane firmy (5)

	LEG	ALM	HEI	SIM	DAS
Odchylenie standardowe	22	25	11	5	1

Tabela 39: Odchylenie standardowe punktów za dane firmy (5)

Wyniki są bardzo niespodziewane. Możliwe, że "0", za dane miesiące zaburzają wygląd punktowy, z tego względu też sprawdzimy od razu jak to ma się do "wyciętych" wyników.

	AMP	ALH	PRU	ALL	SID
Odchylenie standardowe	211	121	217	116	50

Tabela 40: Odchylenie standardowe punktów za dane firmy (5)

	WAR	NNN	CMM	GEN	GEM
Odchylenie standardowe	39	646	38	59	35

Tabela 41: Odchylenie standardowe punktów za dane firmy (5)

	LEG	ALM	HEI	SIM	DAS
Odchylenie standardowe	20	32	12	10	

Tabela 42: Odchylenie standardowe punktów za dane firmy (5)

Jak widać tabele 40-42 są dużo bardziej miarodajne. Czemu? Przyjźyjmy się firmie NNN z tabeli 38 oraz 41. Wyniki BARDZO się od siebie różnią. Jest to spowodowane najwyższym wynikiem, o którym dużo wcześniej wspominałem. Zwykła osoba może naciąć się na ten błąd. Proszę spojrzeć - AMP w obu przypadkach ma powyżej 200 punktów. Jeśli dla pierwszego przypadku weźmiemy odchylenie standardowe można powiedzieć, iż NNN jest bardziej stałą firmą punktowo. Z wcześniejszych jednak analiz wiedzieliśmy, że to nieprawda. Stąd też widząc tego typu wynik musiałem, zrobić według mnie poprawniejsze tabele. Czemu więc nie ma dla firmy DAS wyniku w tabeli 42? Czy jest to błąd? Swego rodzaju. Mamy tylko jedną daną z tej firmy, przez co nie jesteśmy za bardzo w stanie obliczyć dla tej firmy odchylenia standardowego. Jeśli chcielibyśmy znaleźć "naj<sup>2</sup>" firmę (najwięcej punktów + stabilność) możemy zrobić od tego kolejną tabelkę i nową podsekcję.

## 1.14 Naj<sup>2</sup>

Jakie będą zasady? Firma dostaje punkty odpowiednio za najniższe odchylenie standardowe (bez firm DAS) - im niższe tym mniej punktów w naszym rankingu. Drugim z kryterium będzie suma zarobków za daną firmę - im więcej zarobków tym mniej punktów. Firma która ma najmniej punktów ma miano naj<sup>2</sup>:

```
## [1] 13
## [1] "^ AMP"
## [1] 13
## [1] "^ ALH"
## [1] 16
## [1] "^ PRU"
## [1] 14
## [1] "^ ALL"
## [1] 13
## [1] "^ SID"
## [1] 13
## [1] "^ WAR"
## [1] 21
## [1] "^ NNN"
## [1] 14
## [1] "^ CMM"
## [1] 18
## [1] "^ GEN"
## [1] 15
## [1] "^ GEM"
## [1] 14
## [1] "^ LEG"
## [1] 16
## [1] "^ ALM"
```

```
## [1] 15
## [1] "^ HEI"
## [1] 15
## [1] "^ SIM"
```

Jak widać naj<sup>2</sup> ma najniższy wynik wielkości równej 13, a są to firmy :

1. AMP
2. ALH
3. SID
4. WAR

### 1.15 Kolejny estymator - współczynnik zmienności

Zacznę od wyjaśnienia - estymator to statystyka, służąca do szacowania wartości parametru rozkładu, w naszym przypadku ilości zdobytych punktów. Po co więc wprowadzać kolejny estymator - współczynnik zmienności. Albowiem warto sprawdzić, czy nasze wyniki, które mówiły nt. stabilności zdobywanych punktów w danej firmie sprawdzają się z inną strategią. Jak się więc liczy współczynnik zmienności? Otóż odchylenie standardowe dzielimy przez średnią arytmetyczną i tyle. Można też pomnożyć to razy 100% by mieć to przedstawione procentowo. Zobaczmy jak ten współczynnik ma się do naszych firm:

	AMP	ALH	PRU	ALL	SID
Współczynnik [%]	79	73	92	101	74

Tabela 43: Współczynnik zmienności punktów za dane firmy (5)

	WAR	NNN	CMM	GEN	GEM
Współczynnik [%]	78	90	86	118	88

Tabela 44: Współczynnik zmienności punktów za dane firmy (5)

	LEG	ALM	HEI	SIM	DAS
Współczynnik [%]	62	84	71	91	

Tabela 45: Współczynnik zmienności punktów za dane firmy (5)

Po wynikach widać, że warto było to sprawdzić. Firma która posiadała najniższe odchylenie, ma teraz jedno z wyższych współczynników zmienności. To pokazuje, że warto spradzać nasze dane na różne sposoby.

## 1.16 Procentowy udział firm

Na koniec sprawdzimy jak firmy miały wpływ na na zyski:

	AMP	ALH	PRU	ALL	SID
Udział [%]	22.01	16.58	15.18	10.28	6.34

Tabela 46: Udział danych firm (5)

	WAR	NNN	CMM	GEN	GEM
Udział [%]	5.24	5.13	5.03	4.89	4.13

Tabela 47: Udział danych firm (5)

	LEG	ALM	HEI	SIM	DAS
Udział [%]	2.33	1.65	0.92	0.24	0.03

Tabela 48: Udział danych firm (5)

Jak widać, 3 pierwsze firmy łącznie mają udział w zdobytych punktach na poziomie  $> 50\%$ , co pokazuje jak bardzo są one ważne i dobre.

## 2 Podsumowanie

Teraz postaram się zebrać jak najwięcej informacji zdobytych podczas analizy punktowej:

- MetLife zdobyła najwięcej punktów a najmniej - DAS;
- Razem zdobyto 27 871 punktów - sporo;
- Najbardziej produktywnym miesiącem był marzec 2020 - początek pandemii;
- Najwyższy wynik był wysokości 1172 punktów i był spowodowany grupówką w NNN;
- Średnia nie zawsze jest najlepsza do sprawdzenia średniej;
- Dzięki wykresom można zauważyć spory rozwój;
- Odchylenie standardowe oraz współczynniki zmienności pomogły w znalezieniu najstabilniejszych firm;
- Warto sprawdzać za pomocą różnych estymatorów nasze wyniki - niekoniecznie będą te same;
- Najstabilniejszą firmą okazała się ALH - przyrównanie  $naj^2$  do współczynnika zmienności;
- Firmy AMP, ALH oraz PRU stanowią 53% zdobytych wszystkich punktów;