	Politechnika Bydgoska im. J. J. Śniadeckich Wydział Telekomunikacji, Informatyki i Elektrotechniki Zakład Systemów Teleinformatycznych		
Przedmiot	ARiSC		
Prowadzący	prof. dr hab. inż. prof. PBŚ Piotr Cofta		
Temat	<i>Project Raport 1</i>		
Student	Cezary Tytko		
Ocena		Data oddania spr.	

Etap 1. – wstępna analiza danych

Zacząłem od wczytania danych (dla tygodnia 37 i 38), wczytanie nie spowodowało żadnych błędów, zatem zakładałem że jak na razie są one przynajmniej poprawnie sformatowane, zakres czasów też się zgadza, jak i liczność pomiarów jest zgodna z oczekiwaną czyli 604800 ($7 * 24 * 60 * 60$), danych nam nie brakuje, ani nie jest ich nadto, należy jeszcze sprawdzić czy są one odpowiednio rozmieszczone.

Aby sprawdzić rozmieszczenie, obliczyłem różnicę kolejnych czasów, wiedziałem że wartości te nie będą idealnie równe, ale powinny być bardzo blisko 1s i w istocie tak było, dla pierwszego tygodnia największy odstęp jaki odczytałem wyniósł 1.174 ms, a najmniejszy (należałoby się spodziewać że średnia z tych dwóch wartości powinna być w okolicy 1s) wyniósł 814 ms, tak odległych wartości od zakładanej było bardzo niewiele, a i tak uważam że taki odstęp jest do przyjęcia, zatem dane dla pierwszego tygodnia są poprawne jeżeli chodzi o ich rozmieszczenie w czasie. Dla drugiego tygodnia największa różnica wynosi 2.230 ms a najmniejsza 123ms, jednak jest to pojedynczy przypadek, i pomimo odstępu ponad 2 sekundy liczność danych się zgadza oraz taka sama anomalia występuje w tym samym momencie dla wszystkich morzonych parametrów, zakładałem (nie wiem tego na pewno) że mogło być spowodowane opóźnieniem programu/bazy danych, a

pomiary zostały zakolejkowane, dlatego po długiej przerwie kolejne zapisy następowały z przyśpieszeniem i aż wyrównały do interwału 1 s

Aby dostosować dane do przyszłej analizy zdecydowałem zresetować index czasu i założyć stały interwał 1 s dla wszystkich pomiarów, aby wyeliminować wartości z poziomu ms, oraz naprawić rozłożenie danych gdzie przerwa wynosiła 2s, a następnie 0,1 s. W takim przypadku zwykłe zaokrąglenie, od 0,5 , w górę, czy w dół nie zadziała nie tylko ze względu na różnicę ponad 2 s, ale nawet przy prawie stałym interwale 1 s możemy znaleźć się w punkcie gdzie z naszej perspektywy nieistotna różnica sprawi że w wyniku zaokrąglenia dwa pomiary otrzymają identyczny czas, oraz w parze do tego powstanie gdzieś luka, dlatego lepiej wygenerować sztucznie indeksy i je przypisać w kolejności jakiej dane były zapisane.

Podstawowe parametry:

1. Tydzień

Description:			
	f1	thdv1	v1
count	604800.000000	604800.000000	604800.000000
mean	50.007689	1.249792	228.602391
std	0.021855	0.440625	2.269817
min	49.867226	0.000000	222.466995
25%	49.993290	0.981549	226.572998
50%	50.008327	1.262457	228.815994
75%	50.022324	1.515106	230.524002
max	50.115768	6.100586	234.423004

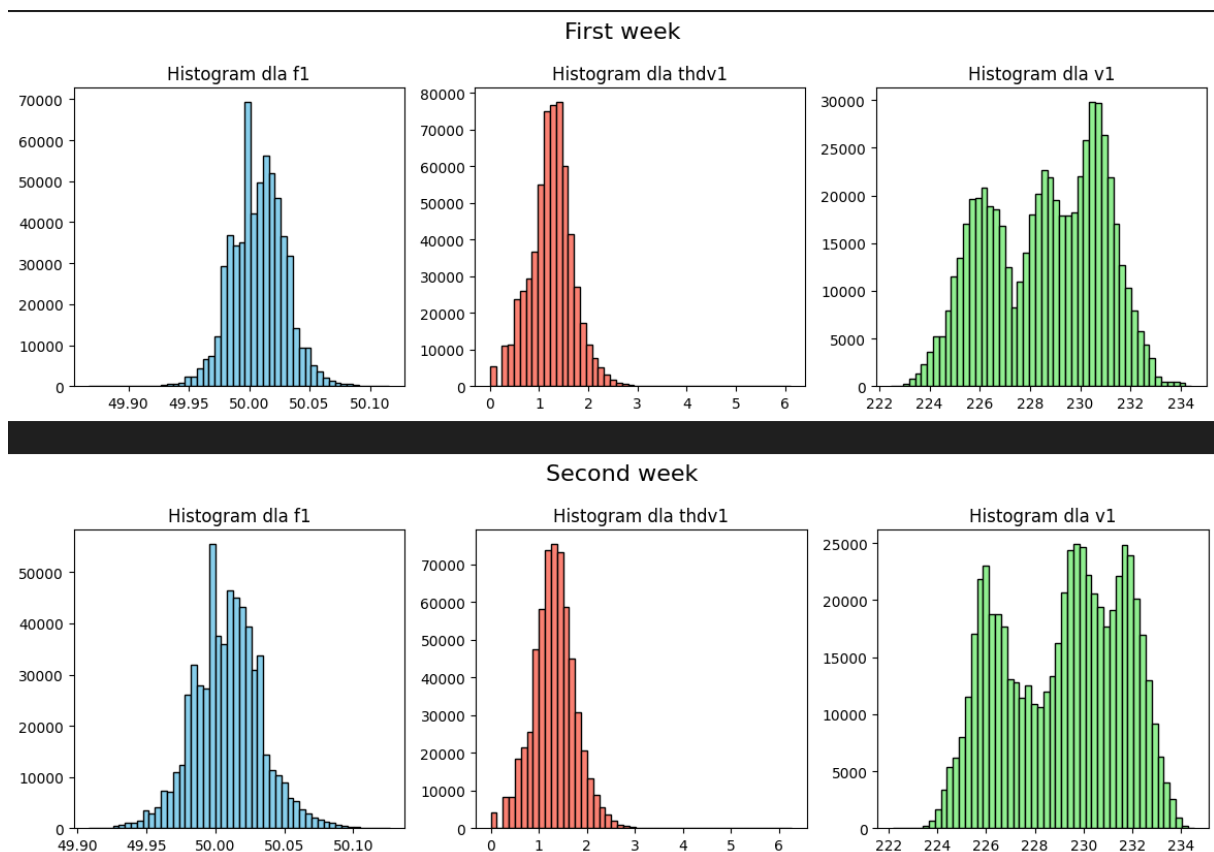
Wyliczone parametry zgadzają się s tym czego należało się spodziewać, czyli napięcia w okolicy 230 V z pewnymi odchyłami, oraz średniej częstotliwości 50 Hz z niewielki odstępstwami, thdv1 niewiele mówi nam w prostu, ale może okazać się przydatne w późniejszym etapie, związanym z uczeniem i przewidywaniem wartości.

2. Tydzień

Description:			
	f1	thdv1	v1
count	604800.000000	604800.000000	604800.000000
mean	50.008947	1.312335	229.123174
std	0.024600	0.444818	2.449131
min	49.908211	0.000000	222.171005
25%	49.993065	1.032423	226.819000
50%	50.009426	1.315769	229.485001
75%	50.023880	1.583095	231.225006
max	50.126598	6.265997	234.520996

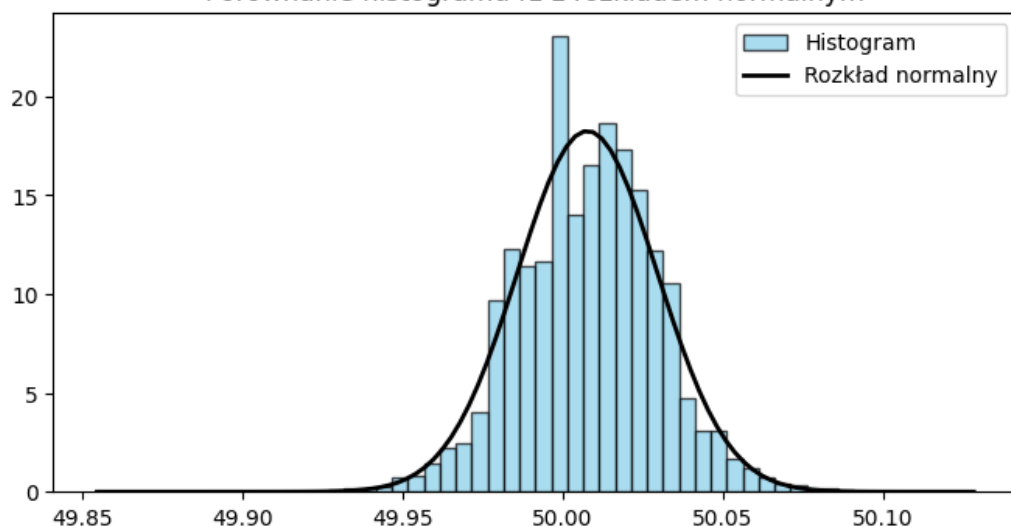
Analogicznie jak dla pierwszego tygodnia dane wyglądają dobrze i nie ma anomalii w postaci np. zaniku napięcia

Histogramy:

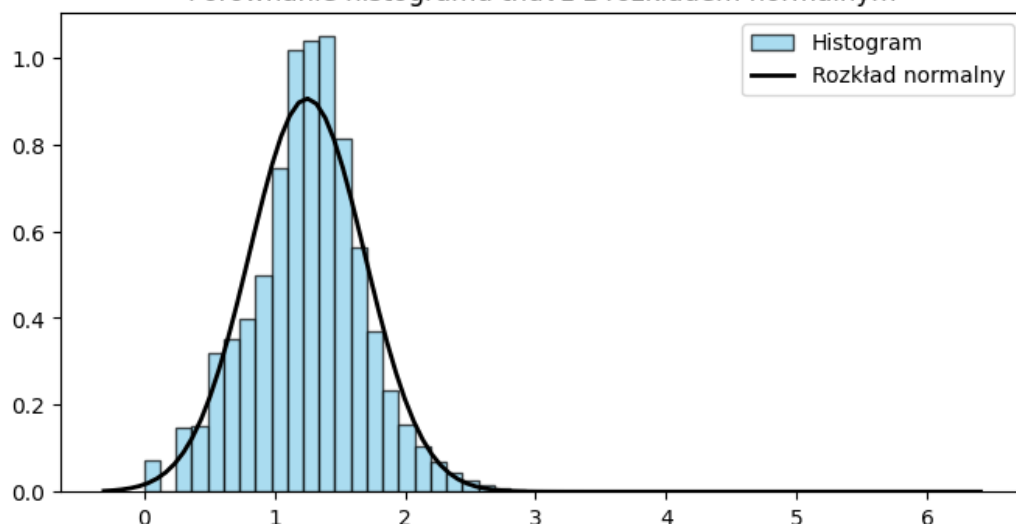


Postanowiłem nałożyć na wykresy hipotetyczny rozkład normalny dla takich danych i wizualnie sprawdzić czy się one pokrywają w znacznym stopniu (dla v1 od razu można stwierdzić że rozkład ten nie jest normalny).

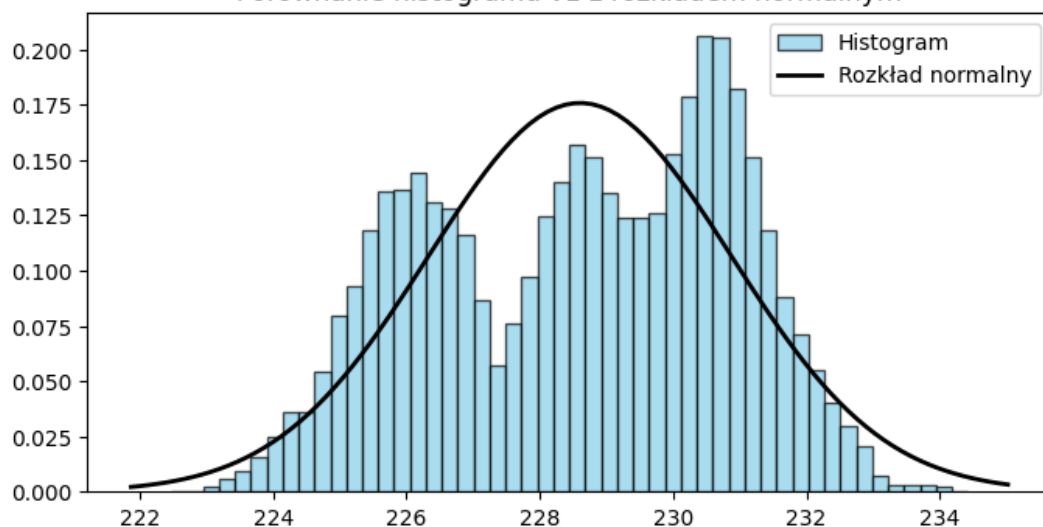
Porównanie histogramu f1 z rozkładem normalnym



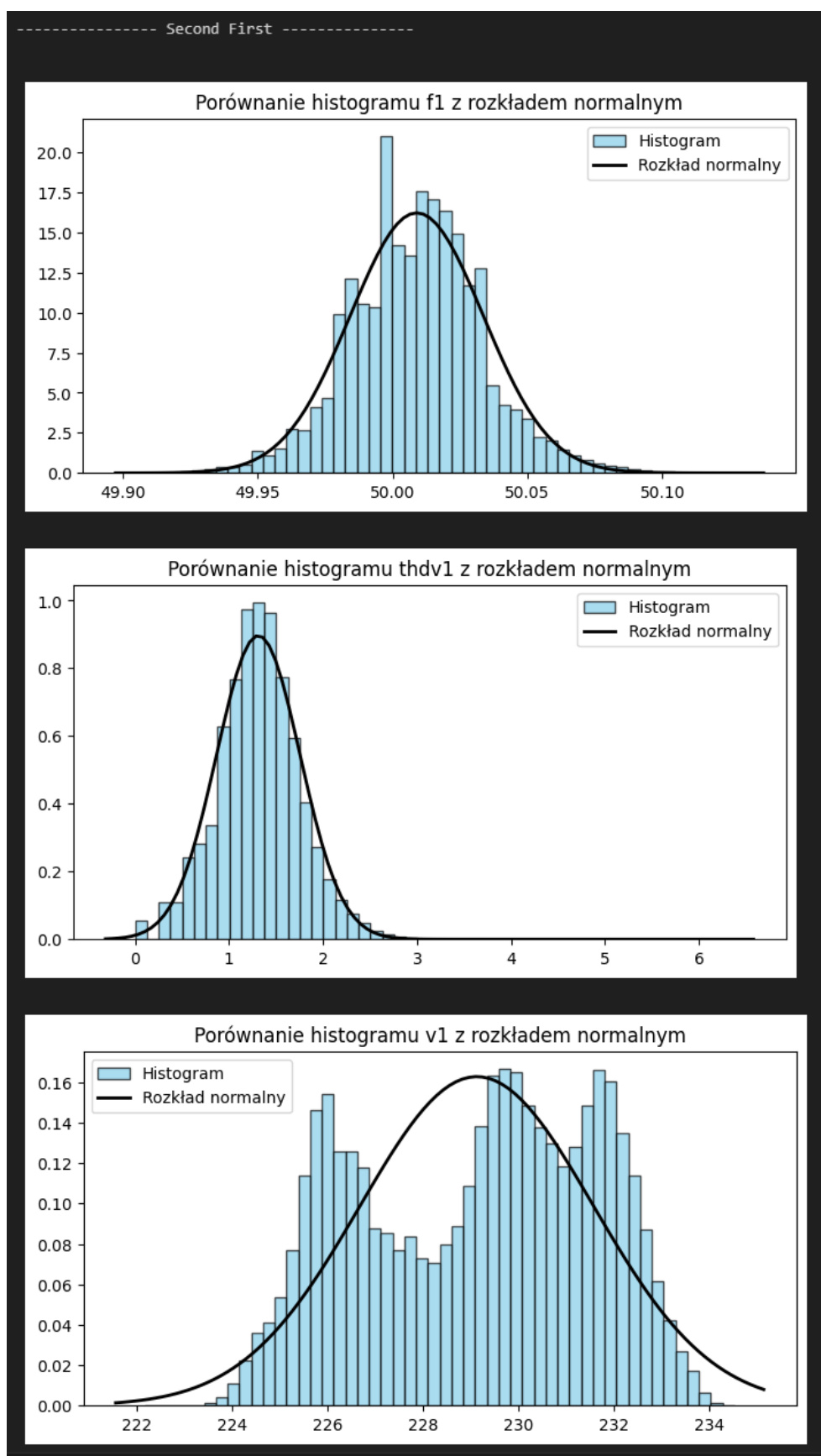
Porównanie histogramu thdv1 z rozkładem normalnym



Porównanie histogramu v1 z rozkładem normalnym



Widać że zdecydowanie wykres $thdv_1$ pokrywa się z rozkładem normalny, to samo można by powiedzieć o wykresie dla f_1 , z wyjątkiem zdecydowanie wyróżniającej się wartości bardzo bliskiej 50 Hz, która to jest oczekiwaną i zadana wartością w sieci elektrycznej, natomiast dla wykresu v_1 widzę połączenie trzech rozkładów normalny, na tym etapie niestety niewiele więcej jestem w stanie powiedzieć, nie wiem jak dokładnie wyglądały warunki badania, mógłbym np. przypuszczać że mamy do czynienia z trzema różnymi stanami w sieci, które mogą być określane przez lokalne obciążenie sieci urządzeniami znacznie oddziaływać na napięcie, jak silniki elektryczne, kondensatory, grzałki.



Takie same wnioski wyciągnąłbym z histogramów dla drugiego tygodnia, jak dla poprzedniego