UNIWERSYTET ZIELONOGÓRSKI

WYDZIAŁ MATEMATYKI, INFORMATYKI I EKONOMETRII

kierunek: Inżynieria danych
specjalność: Modelowanie i analiza danych

NIKODEM SYTNIEJEWSKI

Wpływ zmian pogodowych na kształtowanie się cen wybranych produktów na przestrzeni lat 2013 - 2016

The impact of weather changes on prices of selected products over the years 2013-2016

Promotor pracy inżynierskiej: dr Joachim Syga

....
(podpis promotora)
....
(data)

ZIELONA GÓRA, 2018

Spis treści

1. Wstęp

Deszcz, śnieg, mokro, zimno... zdecydowana większość z nas nie lubi takiej aury. Najlepiej byłoby gdyby dzień zawsze witał nas słońcem, co na pewno korzystnie wpływa na nasze dobre samopoczucie i dzięki czemu z optymizmem podejmujemy codzienną pracę czy naukę.

Nie zawsze jednak pamiętamy - często narzekając na warunki pogodowe ("znowu leje", "ale gorąco"), że owa zmienność warunków atmosferycznych w naszej strefie klimatycznej wymiernie wpływa na otaczającą nas przyrodę i rolnictwo. Tak jest w przypadku wszelkich upraw roślinnych, gdzie, z jednej strony - długo utrzymująca się susza potrafi wyrządzić wiele szkód, a więc wysychanie roślin co znacząco obniża plony.
Z drugiej zaś - stałe opady deszczu i niska temperatura powietrza w okresie wegetacji roślin powoduje ich obumieranie i rozwój chorób grzybowych, co również przekłada się na niższe zbiory. Dzięki odpowiednim warunkom atmosferycznym następuje optymalny rozwój roślin od dziesiątków lat uprawianych w naszym klimacie, są one dobrej jakości i wydają obfite plony.

O tym jak groźna może być susza mówi pan Marek Kubiak, rolnik z województwa łódzkiego: „Od kilku lat mieliśmy suszę i zbierane ziarno miało niewielki procent wilgoci. Ale obecny, 2016 rok, jest inny. Długo było sucho i deszcz nie padał, choć był bardzo potrzebny. Teraz, gdy żniwa tuż-tuż, popaduje i jest chłodno. Nie wiem, czy zmoczone deszczem zboże zdąży wyschnąć, zanim przyjedzie kombajn…”, [5]

Jak widać na przykładzie wypowiedzi pana Marka, nie tylko susza wyrządza wiele szkód podczas uprawy, ale i deszcz, który tak bardzo potrzebny w okresie wzmożonej wegetacji zbóż, nie jest pożądany bezpośrednio przed zbiorem. Summa summarum - wilgotne zboże jest gorszej jakości i rolnikowi trudno jest uzyskać dobrą cenę w skupie.

Przyjrzyjmy się również produktom mlecznym. Producenci mleka nie mają ostatnio powodów do narzekań. Od kilku miesięcy utrzymuje się stosunkowo wysoka cena w skupach białego surowca. Jednak zdaniem analityków, koniunktura może się załamać. „Obecnie światowy rynek mleka znajduje się najprawdopodobniej u szczytu wzrostowej fazy cyklu, która utrzymuje się od II kw. 2016 r. Istnieje wysokie prawdopodobieństwo, że w III kw. 2017 r. światowy rynek osiągnie punkt zwrotny i wejdzie w spadkową fazę cyklu. Wsparcie dla takiego scenariusza stanowią napływające dane wskazujące na stopniową odbudowę podaży mleka wśród najważniejszych eksporterów produktów mlecznych na świecie (m.in. w Nowej Zelandii oraz w UE), przy jednoczesnym osłabieniu globalnego popytu na produkty mleczne. Musimy jednak pamiętać, że wpływ na ceny w naszych krajowych skupach ma wiele czynników. Jednym z ich jest pogoda, zarówno na naszym kontynencie jak i w Nowej Zelandii – największego eksportera produktów mlecznych” - mówi pan Jakub Olipra ekonomista, Departament Analiz Makroekonomicznych, Credit Agricole Bank Polska S.A. [7]

W niniejszej pracy przeanalizujemy wpływ pogody na cenę ziemniaków oraz mleka. Do analizy potrzebne nam będą historyczne dane dotyczące pogody oraz dane GUS dotyczące cen omawianych produktów. Okres, którym będziemy się zajmować, będzie obejmował lata 2013-16. Czynniki pogodowe, które zostały wzięte pod uwagę to: ilość opadów atmosferycznych oraz średnia temperatura.

Te dwa czynniki, a więc częstotliwość opadów i temperatura powietrza najbardziej wpływają na jakość oraz plonowanie upraw. Wiatr lub mgła raczej nie oddziałowują w jakiś znaczący sposób, dlatego nie zostały wzięte pod uwagę.

Praca składa się z następujących rozdziałów:

- rozdział drugi
Opis podstawowych pojęć, które są użyte w pracy.

- rozdział trzeci
Dane pogodowe oraz dane dotyczące cen wybranych produktów.

- rozdział czwarty
Zastosowanie regresji liniowej na wykresach zależności cen produktów od poszczególnych czynników pogody w poszczególnych latach.

- rozdział piąty
Próba przekształceń danych w celu znalezienia innej niż liniowa funkcji, która przybliży nam układ punktów dla wybranych przypadków.

- rozdział szósty
To samo, co w rozdziale czwartym, przy czym badamy cały okres 2013-2016.

- rozdział siódmy
Podsumowanie oraz wnioski.

- rozdział ósmy
Sugestie na przyszłość, dzięki którym otrzymywane wyniki będą lepsze.

- rozdział dziewiąty
Bibliografia. Wszystko, co było potrzebne do pisania pracy.

2. Zanim rozpoczniemy...

Na początku kilka słów wyjaśnienia.

Zanim zajmiemy się dokładniejszą analizą cen wybranych produktów oraz wpływu pogody na ich kształtownie, przytoczone i wyjaśnione zostaną pewne istotne pojęcia, które będą używane w dalszej części pracy.

2.1 Korelacja

Korelacja to miara współwystępowania dwóch zmiennych. Mimo, iż stwierdzenie „miara związku między zmiennymi" brzmi bardzo naukowo i poważnie, tak naprawdę z korelacją w naszym codziennym życiu mamy do czynienia częściej niż nam się wydaje. Któż z nas bowiem nie słyszał narzekań typu: „zawsze kiedy wychodzę z domu bez parasola, zaraz zaczyna padać deszcz". Takie współwystępowanie dwóch zmiennych: brak parasola - opad deszczu to właśnie jest korelacja. Innym książkowym przykładem obrazującym współczynnik korelacji jest stwierdzenie: „im ktoś jest wyższy, tym więcej waży".

Oczywiście, już na pierwszy rzut oka, można zauważyć, iż w obu przywołanych przykładach mamy do czynienia z odmienną skalą pomiarową na której są mierzone zmienne. W pierwszym przypadku (brak parasola - opad deszczu) odwołujemy się do zmiennych nominalnych, w przypadku wagi i wzrostu zmienne są mierzone na skali ilościowej. Dobór współczynnika korelacji zależy właśnie od tego w jakiej skali pomiarowej mierzone są analizowane zmienne.

Najczęściej wykorzystywaną miarą określającą współwystępowanie dwóch zmiennych jest współczynnik korelacji liniowej Pearsona. Za pomocą owego współczynnika jesteśmy w stanie określić zależność pomiędzy zmiennymi mierzonymi na skali ilościowej. To właśnie za pomocą współczynnika Pearsona możemy zweryfikować stwierdzenie: „im ktoś jest wyższy, tym więcej waży". Współczynnik korelacji Pearsona interpretujemy za pomocą dwóch wymiarów: siły i kierunku związku.

Siła związku określa nam stopień istniejącej współzmienności. Współczynnik korelacji może przyjmować wartości od -1 do 1. Jeśli wartość współczynnika zbliża się do 1 lub -1, to mamy do czynienia z silną zależnością. Natomiast korelacja poniżej 0,3 uznawana jest za bardzo słabą lub w ogóle nie istniejącą. Interpretacja taka jest jednak zbyt ogólna i nie możemy jej traktować zbyt ściśle. Na przykład współczynnik równy 0,9 dla socjologów i ekonomistów oznacza silną korelację, a dla fizyków posługujących się wysokiej klasy pomiarami przy badaniu praw przyrody oznacza korelację słabą, [6].

Kierunek korelacji informuje nas zaś o tym, w jaki sposób wartości jednej zmiennej są uporządkowane względem wartości drugiej zmiennej. Możemy mieć do czynienia z korelacją dodatnią - wraz ze wzrostem wartości jednej zmiennej wzrastają wartości drugiej zmiennej, z korelacją ujemną - wraz ze wzrostem wartości jednej zmiennej maleją wartości drugiej zmiennej lub z korelacją równą 0 - brak związku liniowego między analizowanymi zmiennymi, [11].

2.2 Regresja

Regresja jest to metoda statystyczna pozwalająca na badanie związku pomiędzy wielkościami danych. Umożliwia przewidywanie nieznanych wartości jednych wielkości na podstawie znanych wartości innych, [8].

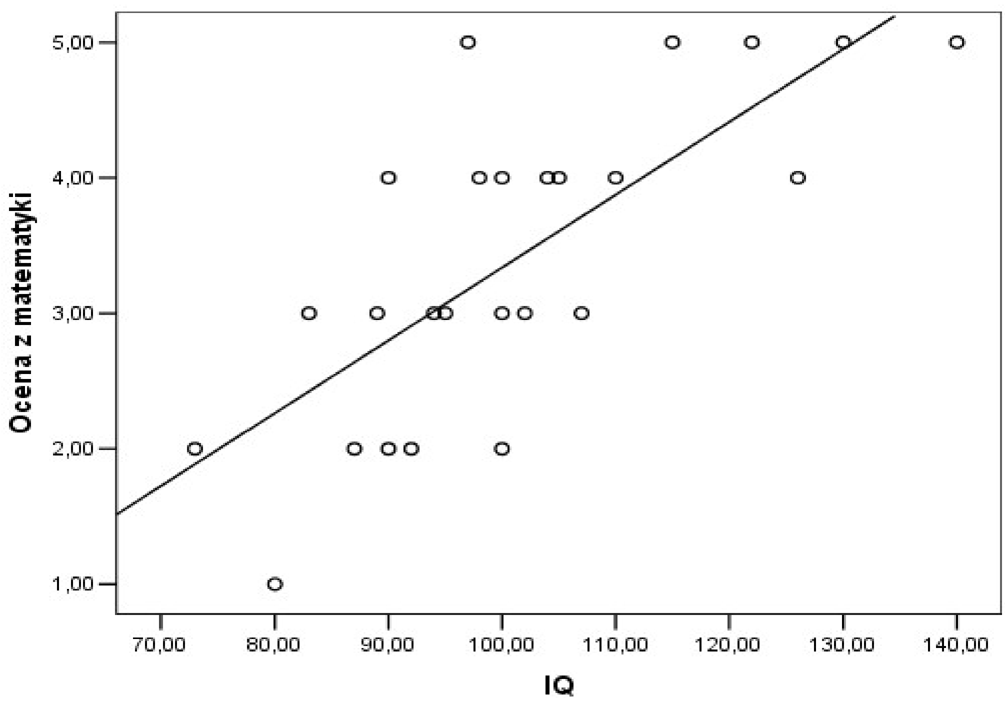
2.2.1 Regresja liniowa

Regresja liniowa jest najprostszym wariantem regresji w statystyce. Zakłada ona, że zależność pomiędzy zmienną objaśnianą a objaśniającą jest zależnością liniową, czyli ***y = ax+b.*** W regresji liniowej, podobnie jak w analizie korelacji, zakłada się, że wzrostowi jednej zmiennej (predyktor) towarzyszy wzrost lub spadek drugiej zmiennej.

Analiza regresji liniowej ma na celu wyliczenie takich współczynników w modelu liniowym, aby ten model jak najlepiej przewidywał wartość zmiennej zależnej i aby błąd tego oszacowania był jak najmniejszy. Przyjrzyjmy się przykładowi.

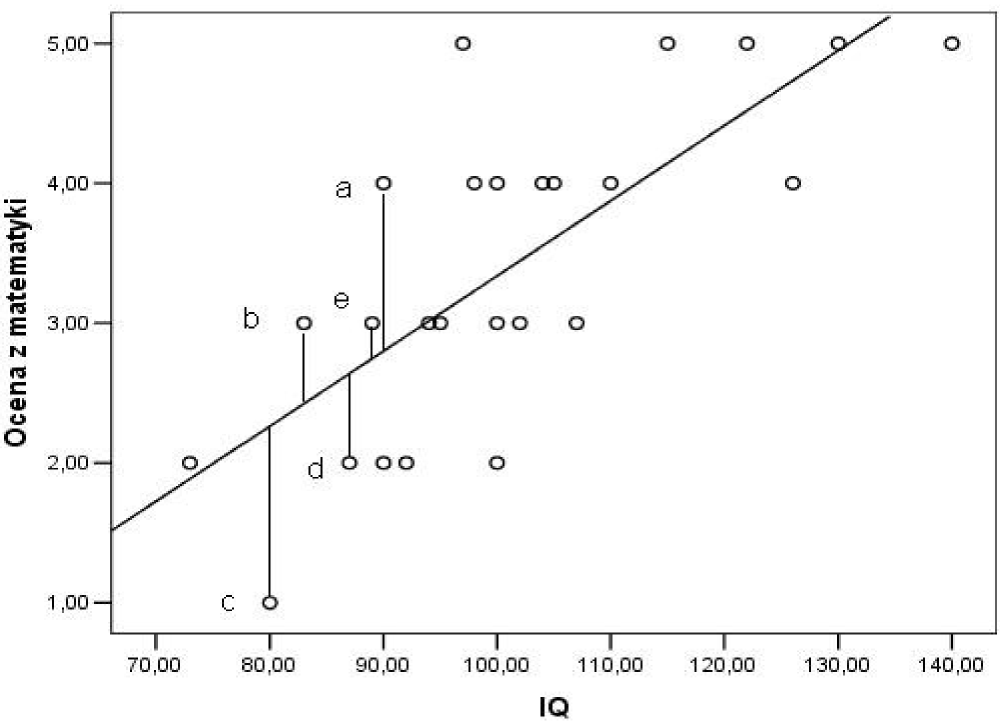
Przykład:

Pewien badacz badał związek pomiędzy poziomem inteligencji uczniów a ich ocenami uzyskiwanymi na koniec roku z matematyki. Na wykresie rozrzutu poniżej przedstawiono uzyskane wyniki:



Jak można zauważyć, zależność pomiędzy IQ a ocenami z matematyki jest liniowa: im wyższy poziom IQ, tym lepsze oceny z matematyki. Oczywiście zależność ta nie jest idealna. Nie zawsze osoby, które mają niższy IQ mają gorsze oceny od osób, które mają wyższe IQ, co widać na wykresie.

Aby wyznaczyć linię regresji, a tym samym wzór w modelu regresji liniowej należy obliczyć współczynniki linii prostej, ***a*** i ***b***, zwane odpowiednio: współczynnikiem regresji oraz wyrazem wolnym. W tym celu wykorzystuje się metodę najmniejszych kwadratów. Metoda ta ma na celu dopasowanie do zebranych danych: pary wyników (poziom IQ oraz oceny z matematyki), takiej linii prostej (model liniowy), która jest do nich najlepiej dopasowana (obliczeniowo). Popatrzymy na wykres rozrzutu na wykresie poniżej:



Aby jak najlepiej zrozumieć metodę najmniejszych kwadratów, będziemy pracować na linii, która już została obliczona. Pogrubiona linia prosta jest linią regresji wyprowadzoną metodą najmniejszych kwadratów. Punkty a, b, c, d, e i pozostałe określają zaobserwowane wartości poziomu IQ i ocenę z matematyki u poszczególnych badanych osób. Linia pionowa (cieńsze proste linie) pomiędzy linią regresji a punktem stanowi błąd oszacowania naszego modelu. Nasz model zakłada, że gdy osoba ma IQ około 90 pkt to jego średnia ocena z matematyki powinna wynosić około 2.8. Jednakże widzimy, że osoby, które mają taki poziom IQ nie zawsze mają oceny na tym poziomie, raz mają lepsze oceny raz gorsze. Różnica pomiędzy oszacowaną linią regresji a faktycznym wynikiem stanowi błąd oszacowania. Dla pozostałych punktów (obserwacji, osób badanych), również możemy wyliczyć błąd tego oszacowania. Metoda najmniejszych kwadratów wyprowadza taką linię prostą, dla której suma kwadratów tych błędów będzie najniższa. Mówiąc prościej... metoda dopasowuje taką linię do zebranych danych, aby ogólny błąd oszacowania (dla wszystkich danych) był jak najmniejszy. Każda inna linia, dostarczałaby większy błąd oszacowania, [13].

Wracając do wzoru na linię prostą, analiza regresji oblicza współczynnik ***a***, zwany współczynnikiem regresji oraz wartość ***b***, zwaną wyrazem wolnym, [9].

2.3 Współczynnik determinacji

Współczynnik determinacji, inaczej zwany współczynnikiem określoności lub *R*-kwadrat, jest miarą tego, jaki procent zmienności zmiennej zależnej (objaśnianej) jest wyjaśniony za pomocą zmiennej niezależnej (czynnik, zmienna objaśniająca, predyktor) bądź modelu statystycznego. Innymi słowy, współczynnik determinacji informuje nas, w jakim stopniu nasz model wyjaśnia zgromadzone dane pomiarowe (zmienną zależną).

Analizowany czynnik może mieć wpływ na zmienną zależną. Pytanie: jak mocny jest jego wpływ? Posłużmy się przykładem.

Przykład:

Pewien badacz chciał sprawdzić, czy poziom inteligencji osób ma wpływ na wykonywanie zadań z matematyki. Wyprowadził model statystyczny, w którym stwierdził, że poziom inteligencji wpływa na ilość poprawnie wykonanych zadań. Ale czy to oznacza, że tylko poziom inteligencji liczy się przy wykonywaniu zadań? Intuicyjnie wiemy, że nie tylko on ma wpływ, wpływać może również poziom wiedzy, sumienność uczenia się, motywacja. Poziom inteligencji jest tutaj jedynie jednym z czynników, którego wpływ nasz badacz chciał sprawdzić. Tak więc cała zmienność wyników (wariancja) z wykonywaniu zadań nie jest wyjaśniona jedynie za pomocą poziomu inteligencji. Dlatego właśnie badacz oblicza współczynnik determinacji, aby określić procent wyjaśnionej zmienności, wariancji uzyskanych wyników. Pozostała część wariancji pozostaje niewyjaśniona poprzez ten czynnik.

Współczynnik determinacji często określa się terminem *R*-kwadrat, (*R*-kwadrat, R^2) z tego powodu, że symbolem współczynnika determinacji jest właśnie R^2. Jak można zauważyć, w przypadku badania zależności liniowej zmiennych, R^2 otrzymujemy poprzez podniesienie do kwadratu współczynnika korelacji liniowej Pearsona *r*. Zatem, jeżeli znany jest współczynnik korelacji liniowej pomiędzy zmiennymi (wyjaśnianą i wyjaśniającą), to podniesienie go do kwadratu da nam wynik współczynnika determinacji.

Inny sposób obliczania współczynnika korelacji wywodzi się z analizy regresji. W modelu sprawdzamy czy dany predyktor bądź grupa predyktorów, czynników, mają wpływ na zmienną zależną, wyjaśnianą. Dla modelu liniowego z jednym predyktorem używamy następującego wzoru:

$$R^2 = \frac{SS_M}{SS_T} = \frac{\sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

gdzie:
R^2 - współczynnik determinacji, R-kwadrat, procent wyjaśnionej zmienności przez model,
SS_M - suma kwadratów dla modelu,
SS_T - suma kwadratów całkowita,
y_t - rzeczywista wartość zmiennej zależnej (zmierzona),
$\hat{y}_t$ - przewidywana wartość zmiennej zależnej (na podstawie modelu regresji),
$\bar{y}$ - średnia wartość rzeczywistej zmiennej zależnej.

Współczynnik determinacji przyjmuje wartości pomiędzy 0 a 1, jednakże najbardziej popularną formą prezentacji tego współczynnika jest jego procentowa postać. Uzyskany wynik mnożymy przez 100% i uzyskujemy procent wyjaśnionej wariancji. Należy zaznaczyć, że jeżeli mamy do czynienia z większą liczbą predyktorów w modelu, to współczynnik determinacji powinien zostać skorygowany.

Zastosowanie współczynnika determinacji w analizach:

- współczynnik determinacji, R-kwadrat, daje nam informację, na ile nasze badanie (nasz założony czynnik) wyjaśnia to, co chcemy mierzyć,

- służy określeniu, na ile poszczególne modele statystyczne, czynniki „dobrze” wyjaśniają to, co chcemy wyjaśnić, która ze zmiennych, (jeżeli badamy kilka), lepiej wyjaśnia zmienną zależną,

- pozwala oszacować, który z analizowanych modeli jest lepszy,

- najczęściej stosowany w modelowaniu statystycznym, ekonometrycznym, niż w zwykłej analizie korelacji, [12].

2.4 Wykorzystane funkcje w programie R

Do otrzymania wyników zamieszczonych w pracy wykorzystano następujące funkcje w programie R:

- *library,*
- *read.table,*
- *plot,*
- *par,*

- *lm,*
- *seq,*
- *lines,*
- *cor,*
- *summary.*

3. Dane

	Miesiac	Rok	Temp_Dzien	Temp_Noc	Opady	Cena_ziemniakow	Cena_mleka
1	styczeN	2013	-0.55	-4.83	70.3	52.12	130.94
2	luty	2013	0.97	-2.41	33.2	56.88	129.45
3	marzec	2013	0.83	-6.34	34.2	48.89	132.46
4	kwiecien	2013	12.28	4.41	23.6	67.38	133.69
5	maj	2013	18.07	9.59	63.9	40.03	134.96
6	czerwiec	2013	21.45	13.00	104.4	41.65	137.57
7	lipiec	2013	25.48	15.31	68.0	32.22	137.65
8	sierpieN	2013	24.03	14.31	42.6	21.32	141.36
9	wrzesien	2013	16.45	9.10	55.3	23.65	144.64
10	pazdziernik	2013	14.86	5.86	18.8	25.45	149.38
11	listopad	2013	6.86	1.66	52.1	22.87	158.28
12	grudzien	2013	5.59	0.03	24.6	46.47	164.89
13	styczeN	2014	1.62	-3.52	36.6	68.73	157.16
14	luty	2014	7.59	-1.34	6.9	85.47	156.72
15	marzec	2014	11.45	1.31	43.9	80.03	156.41
16	kwiecien	2014	14.97	6.21	36.6	76.38	151.44
17	maj	2014	17.28	8.93	114.1	66.07	146.09
18	czerwiec	2014	21.38	12.10	27.3	38.01	145.21
19	lipiec	2014	26.45	16.34	73.6	27.27	139.07
20	sierpieN	2014	21.79	12.90	96.8	21.27	138.43
21	wrzesien	2014	19.24	11.24	52.0	19.09	126.94
22	pazdziernik	2014	15.21	7.69	33.1	23.67	126.56
23	listopad	2014	8.72	3.52	5.5	20.54	132.82
24	grudzien	2014	4.21	-0.41	23.4	34.17	133.91
25	styczeN	2015	3.87	-0.71	63.0	44.19	122.18
26	luty	2015	4.69	-1.45	5.2	43.83	120.74
27	marzec	2015	9.58	2.35	34.0	52.85	121.02
28	kwiecien	2015	13.17	5.00	26.2	54.10	116.43
29	maj	2015	17.94	9.19	18.9	53.18	112.52
30	czerwiec	2015	20.93	12.27	70.5	62.31	110.40
31	lipiec	2015	24.35	16.03	99.6	50.43	108.47
32	sierpieN	2015	28.19	18.84	6.9	44.18	107.81
33	wrzesien	2015	19.50	12.30	45.0	35.27	108.79
34	pazdziernik	2015	12.32	6.48	31.7	30.82	112.03
35	listopad	2015	9.50	5.63	61.3	31.01	113.20
36	grudzien	2015	7.94	4.65	17.6	46.81	112.89
37	styczeN	2016	1.13	-1.94	44.3	58.09	111.80
38	luty	2016	5.76	1.83	50.5	57.20	109.54
39	marzec	2016	7.13	1.39	39.7	66.84	105.89
40	kwiecien	2016	13.10	4.77	52.0	64.93	102.80
41	maj	2016	19.74	10.52	34.8	67.29	100.19
42	czerwiec	2016	22.97	14.13	71.7	68.77	99.97
43	lipiec	2016	23.48	15.23	98.6	51.62	101.28

Obecnie mamy szereg możliwości pozyskania informacji pogodowych. Synoptycy w telewizji czy radiu na bieżąco informują jakiej pogody można oczekiwać w najbliższych dniach. Brane są pod uwagę również potrzeby poszczególnych grup zawodowych jak np. prognozy pogody dla rolników (uwzględnia się temperaturę przy gruncie). Internet czy też aplikacje pogodowe zainstalowane w smartfonach umożliwiają nam nie tylko śledzenie aktualnych tendencji pogodowych, ale dzięki nim mamy również dostęp do długoterminowej prognozy, np. 16-dniowej, kiedy interesuje nas jaka będzie aura w dłuższej perspektywie czasowej.

Istnieją również dane obserwacji meteorologicznych, z których dowiemy się, jaka była pogoda dzień, miesiąc czy nawet kilka lat temu. Ogólne średnie dane pogodowe dla poszczególnych lat czy miesięcy nie były dostępne. Wszystkie historyczne dane dotyczyły poszczególnych dni. Aby uśrednić i uzyskać szczegółowe dane miesięczne, należało najpierw z każdego miesiąca uśrednić temperaturę oraz ilość opadów.

Dane dotyczące temperatury: [4] oraz [1].
Temperaturę będziemy określać w ℃

Dane dotyczące opadów: [2].
Łączna suma wszystkich opadów będzie podawana w mm.

Jeżeli chodzi o dane dotyczące cen produktów (pozycja [3]), to najlepszym źródłem były dane z GUS-u (Główny Urząd Statystyczny). Na interetowej stronie GUS-u można uzyskać dużo szczegółowych informacji o cenach bardzo wielu produktów spożywczych i nie tylko. W porównaniu z danymi pogodowymi, nie wymagały one specjalnych obróbek, gdyż zawierały wszystkie interesujące nas elementy składowe, a więc zarówno dane miesięczne jak i roczne.

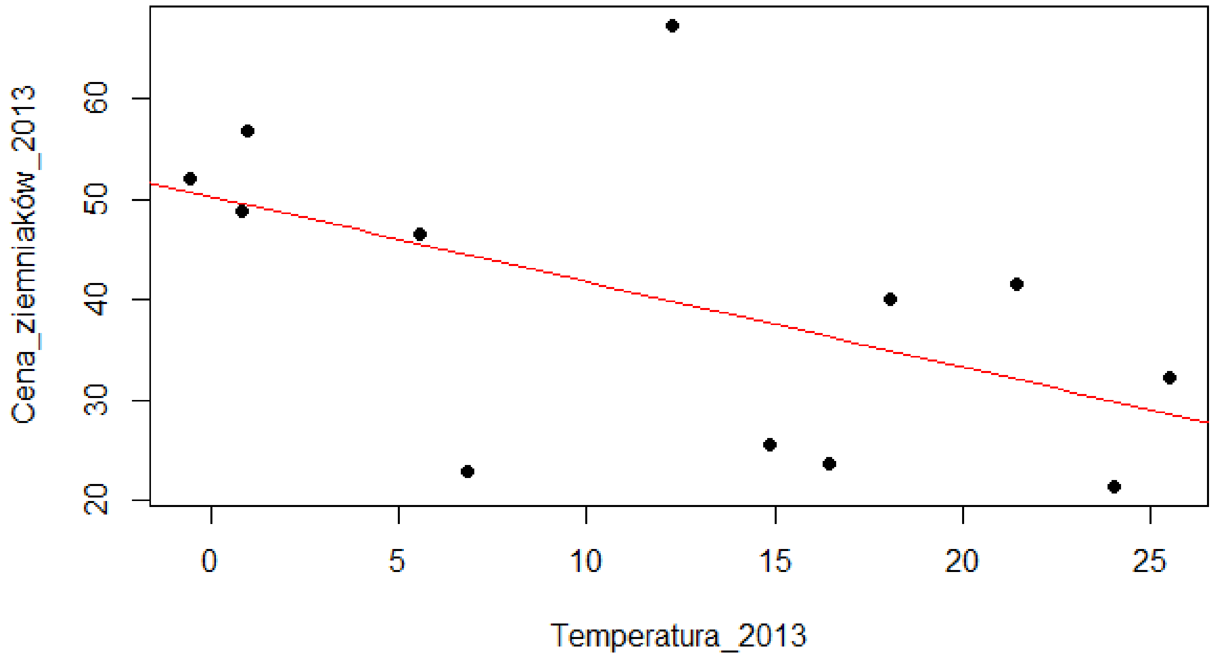
4. Zależności cen poszczególnych produktów od pogody

W niniejszym rozdziale spróbujemy zastosować regresję liniową dla wykresów pokazujących zależność cen poszczególnych produktów od poszczególnych czynników pogody dla określonego roku. Dla każdego wykresu sprawdzimy, ile wynosi korelacja oraz spróbujemy wpasować funkcję liniową do punktów znajdujących się na wykresie.

Bardziej szczegółową analizę dołączamy do modeli, dla których moduł korelacji przekracza wartość 0.5.

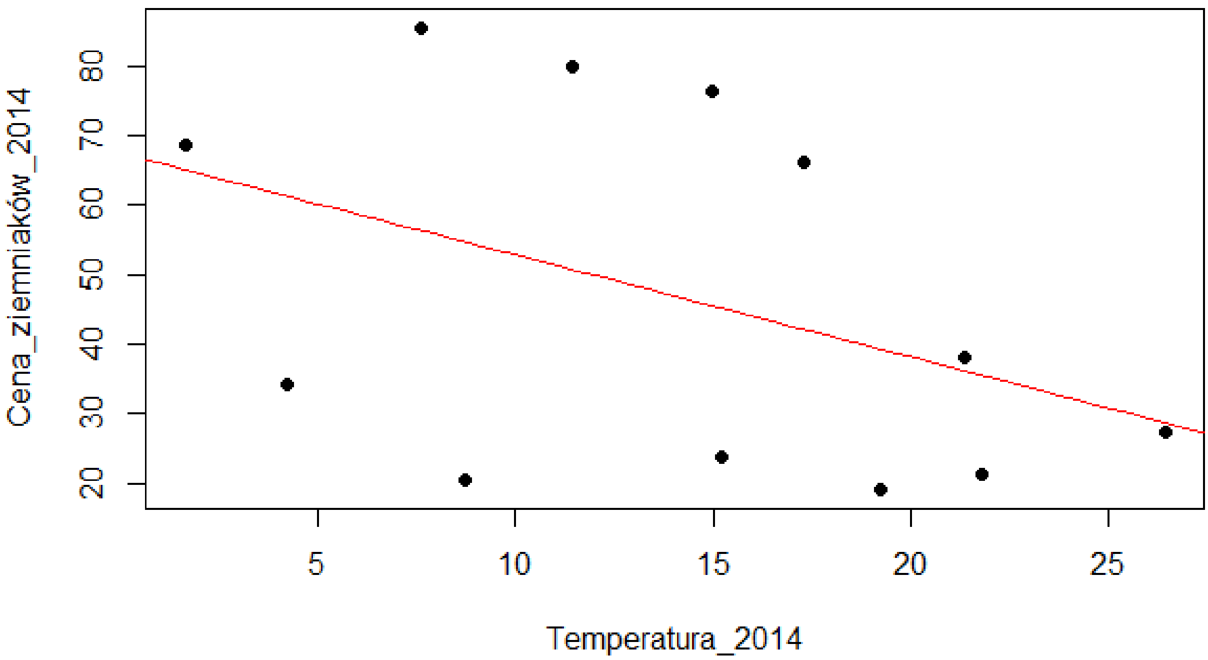
4.1 Zależność ceny ziemniaków od temperatury

a) 2013 rok



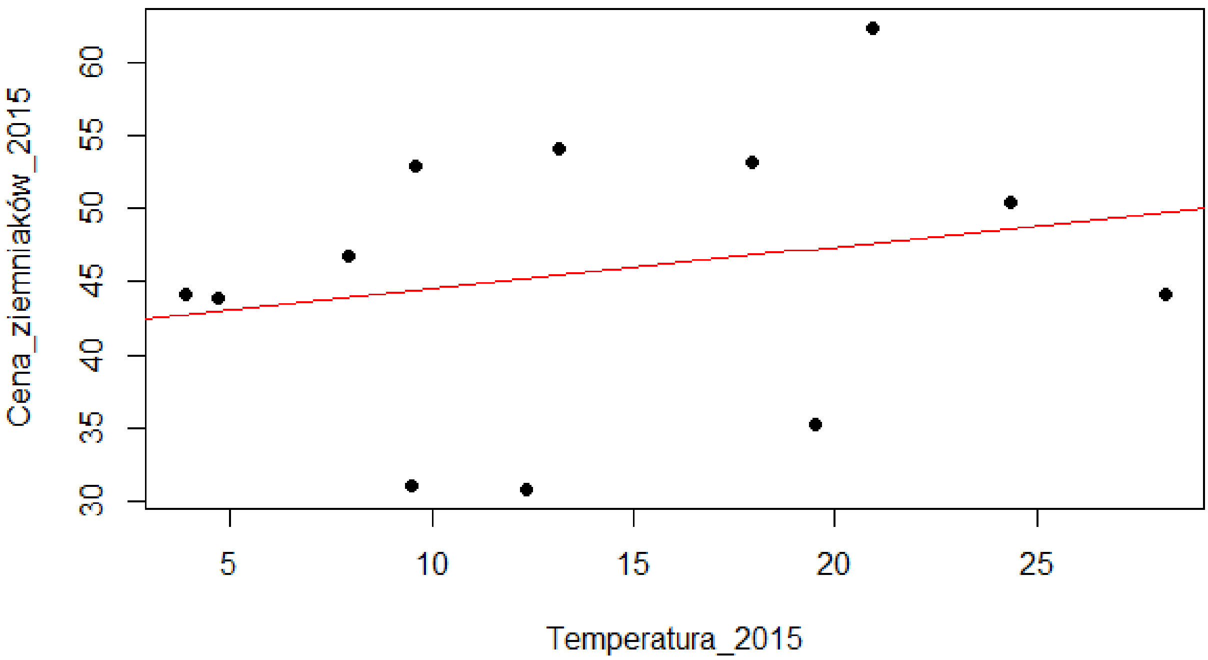
- korelacja umiarkowana równa -0.527
- wzór funkcji: ***y = -0.8526*x + 50.3066***
- przy wyższej temperaturze cena ziemniaków spadała.
  Gdy temperatura wzrośnie o 1℃ to cena ziemniaków spadnie o około 85 gr
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $1.7*10^{-5}$ przy ***x*** oraz 0.0781 przy wyrazie wolnym.
  Zakładając poziom istotności alfa = 0.05, pierwszy ze współczynników ma prawdopodobieństwo błędu mniejsze niż alfa, drugi z kolei ma prawdopodobieństwo błędu większe niż alfa
- *R*-squared = 0.2781. Z tego wynika, że w modelu w około 28% temperatura wyjaśnia cenę ziemniaków
- *p*-value = 0.07805

b) 2014 rok



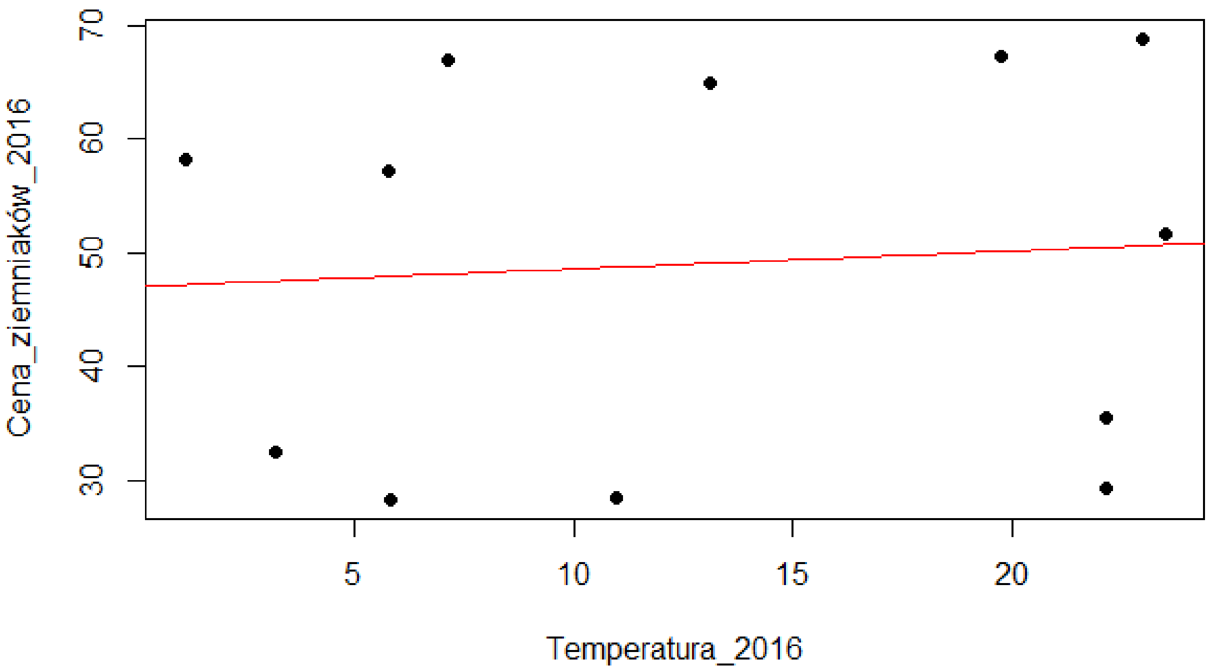
- korelacja umiarkowana równa -0.424
- wzór funkcji: ***y = -1.4642*x + 67.4572***
- zauważamy, że im wyższa temperatura, tym cena ziemniaków jest niższa. Gdy temperatura wzrośnie o 1℃ to cena ziemniaków spadnie o około 1.46 zł

c) 2015 rok



- korelacja słaba równa 0.232
- wzór funkcji: ***y = 0.2878*x + 41.6234***
- brak wyraźnych powiązań pomiędzy temperaturą a ceną ziemniaków, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem temperatury o 1℃ cena ziemniaków wzrośnie o około 28 gr

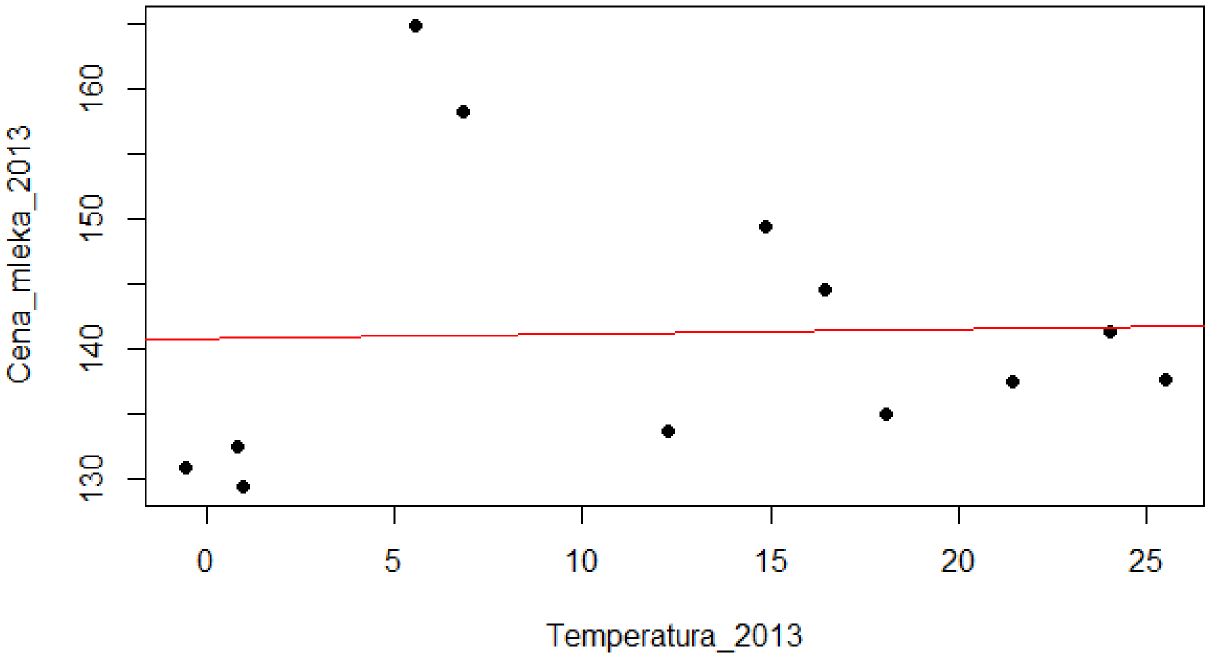
d) 2016 rok



- znikoma korelacja równa 0.0797
- temperatura nie wpływa w żaden sposób na cenę ziemniaków
- wzór funkcji: ***y = 0.1582*x + 46.9785***
- wpływ temperatury na cenę ziemniaków nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem temperatury o 1℃ cena ziemniaków wzrośnie o około 16 gr

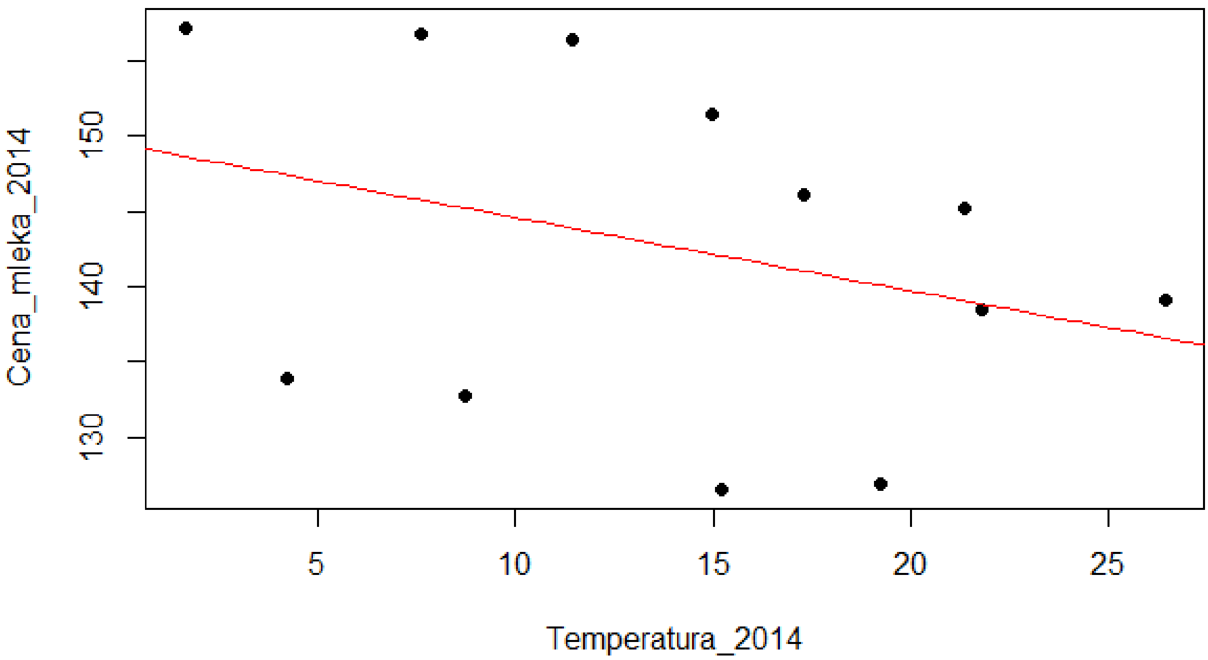
4.2 Zależność ceny mleka od temperatury

a) 2013 rok



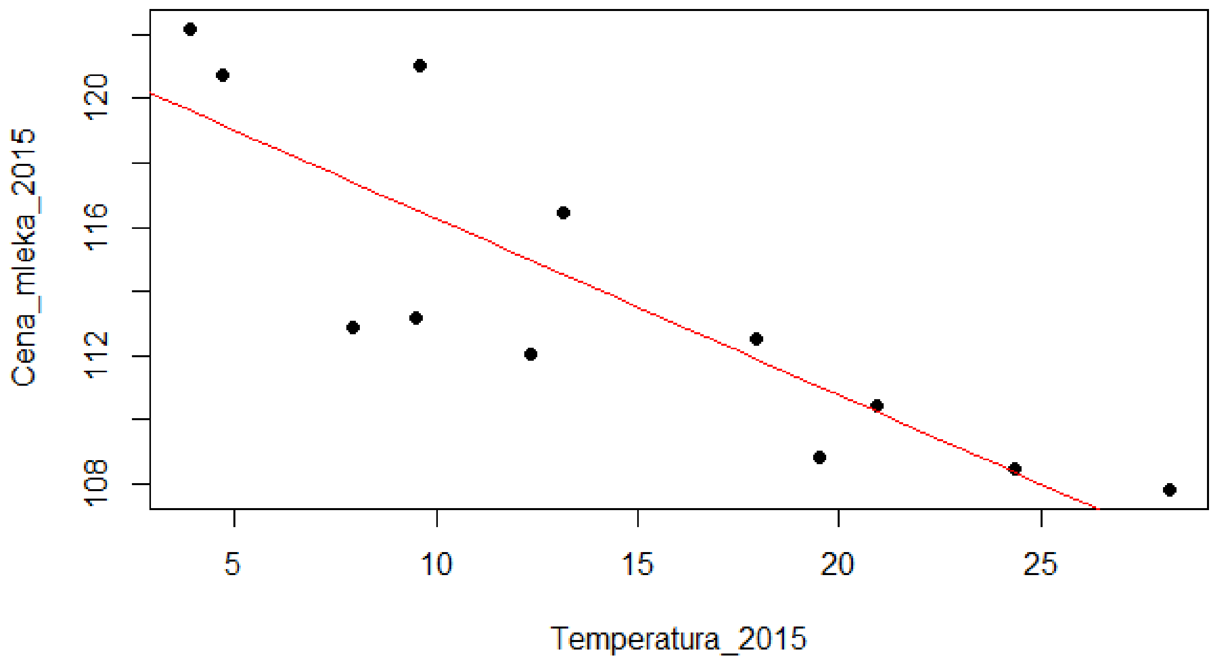
- znikoma korelacja równa 0.0303
- rozkład punktów na wykresie zbliżony jest do wykresu gęstości rozkładu normalnego
- wzór funkcji: ***y = 0.03643*x + 140.82828***
- wpływ temperatury na cenę ziemniaków nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem temperatury o 1℃ cena mleka wzrośnie o około 4 gr

b) 2014 rok



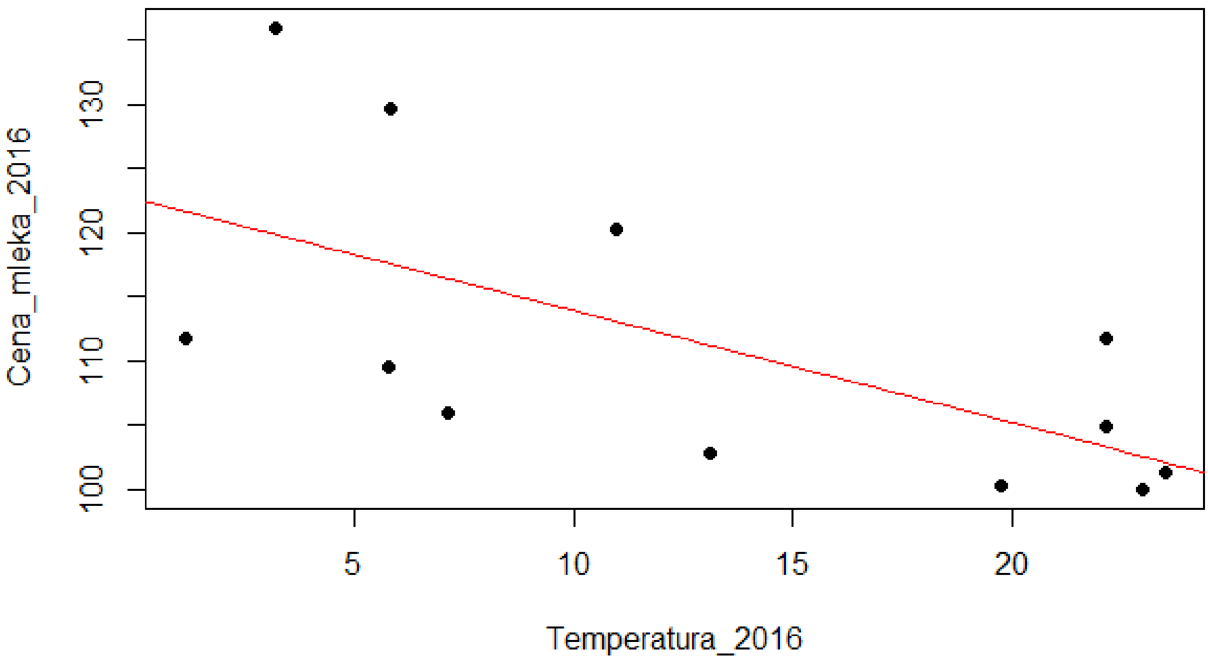
- korelacja słaba równa -0.326
- wzór funkcji: ***y = -0.4847*x + 149.4264***
- wyższa temperatura oznacza niższą cenę mleka.
  Gdy temperatura wzrośnie o 1℃ to cena mleka spadnie o około 48 gr

c) 2015 rok



- bardzo silna korelacja równa -0.845
- wzór funkcji: ***y = -0.5506*x + 121.7638***
- im wyższa temperatura, tym niższa cena mleka.
  Gdy temperatura wzrośnie o 1℃ to cena mleka spadnie o około 55 gr
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $1.07*10^{-14}$ przy ***x*** oraz 0.000533 przy wyrazie wolnym.
  Zakładając poziom istotności alfa = 0.05, zarówno pierwszy jak i drugi współczynnik ma prawdopodobieństwo mniejsze niż alfa
- *R*-squared = 0.7147. Z tego wynika, że w modelu w około 71% temperatura wyjaśnia cenę mleka
- *p*-value = 0.0005335

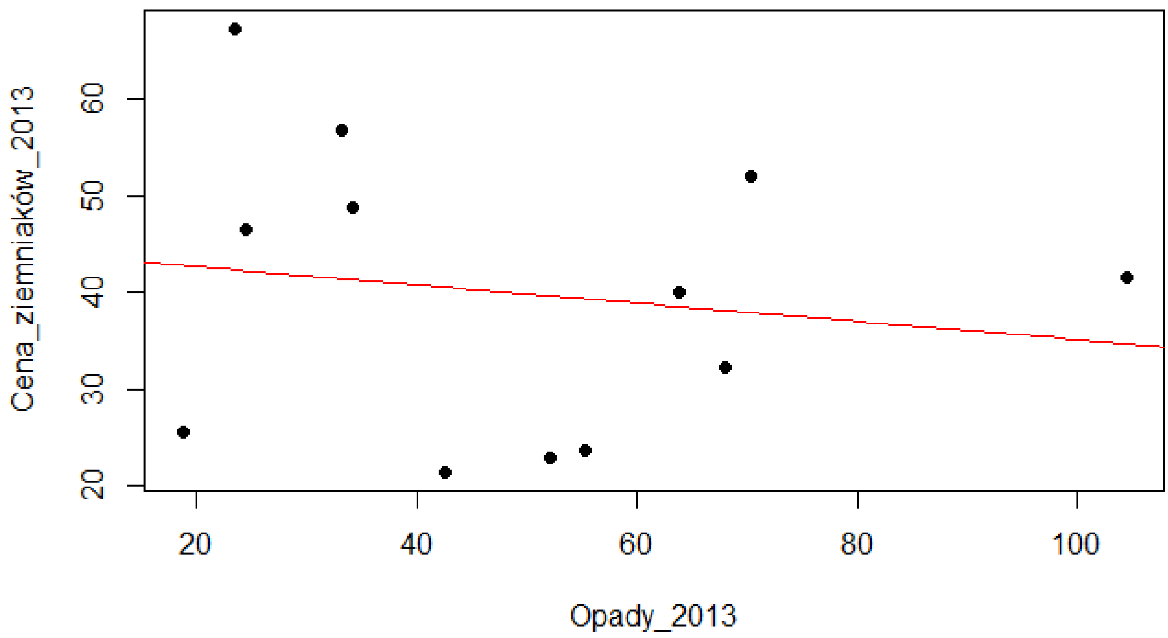
d) 2016 rok



- dość silna korelacja równa -0.634
- wzór funkcji: ***y = -0.8755*x + 122.6710***
- im wyższa temperatura, tym niższa cena mleka.
  Gdy temperatura wzrośnie o 1℃ to cena mleka spadnie o około 88 gr
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $4.41*10^{-10}$ przy ***x*** oraz 0.0268 przy wyrazie wolnym.
  Zakładając poziom istotności alfa = 0.05, zarówno pierwszy jak i drugi współczynnik ma prawdopodobieństwo mniejsze niż alfa
- *R*-squared = 0.4022. Z tego wynika, że w modelu w około 40% temperatura wyjaśnia cenę mleka
- *p*-value = 0.02678

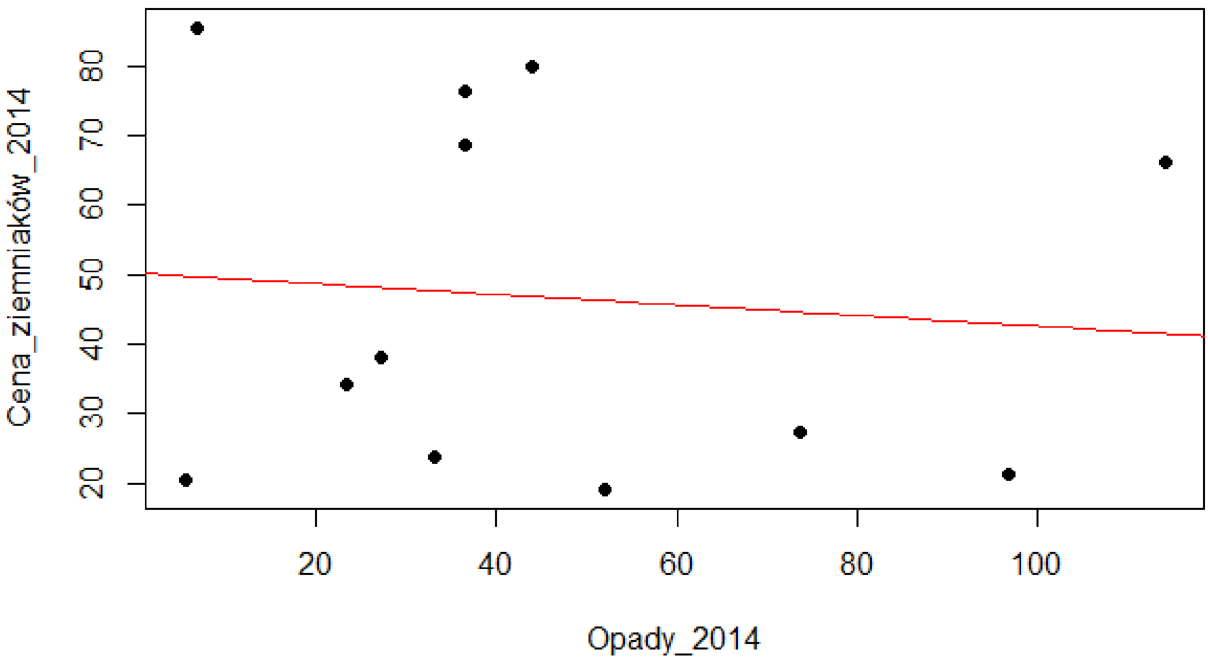
4.3 Zależność ceny ziemniaków od opadów

a) 2013 rok



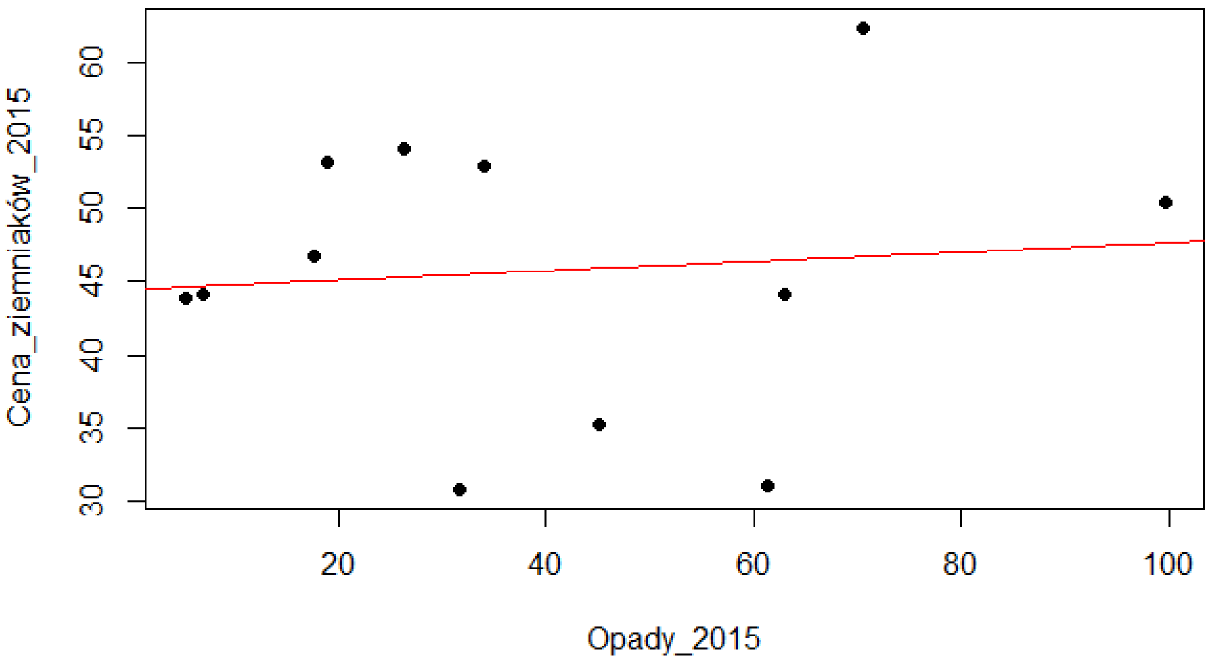
- korelacja słaba równa -0.157
- wzór funkcji: ***y = -0.09481*x + 44.58022***
- wpływ opadów na cenę ziemniaków nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena ziemniaków spadnie o około 9 gr

b) 2014 rok



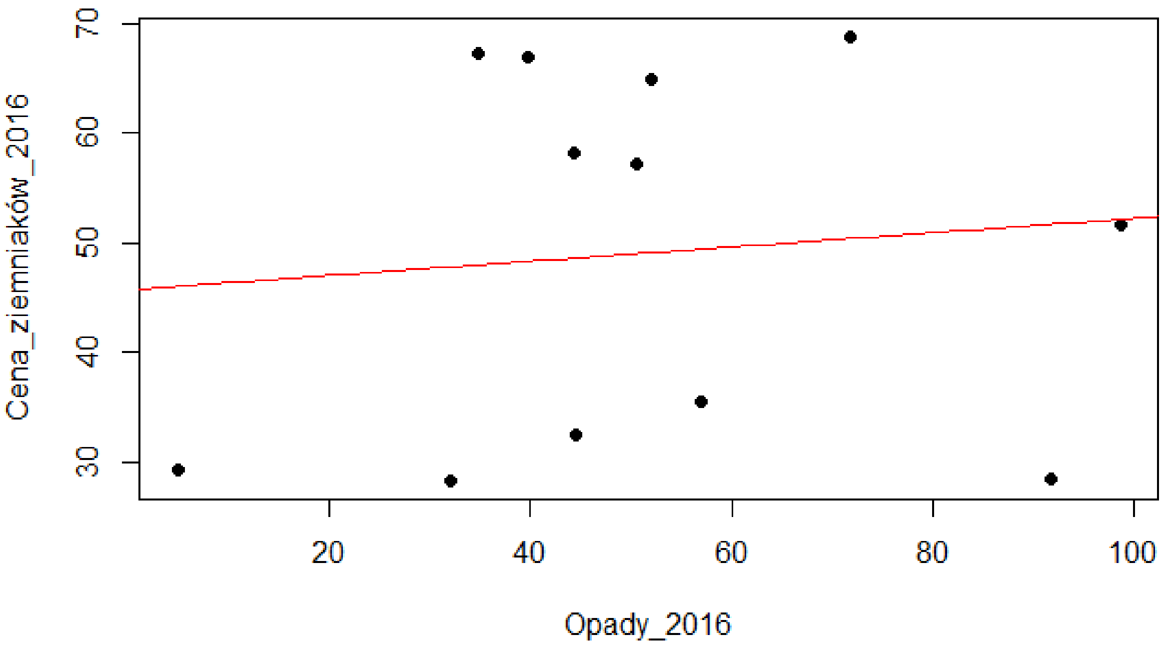
- znikoma korelacja równa -0.0974
- wzór funkcji: ***y = -0.0762*x + 50.2163***
- wpływ opadów na cenę ziemniaków nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena ziemniaków spadnie o około 8 gr

c) 2015 rok



- znikoma korelacja równa 0.0937
- wzór funkcji: ***y = 0.03165*x + 44.48244***
- wpływ opadów na cenę ziemniaków nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena ziemniaków wzrośnie o około 3 gr

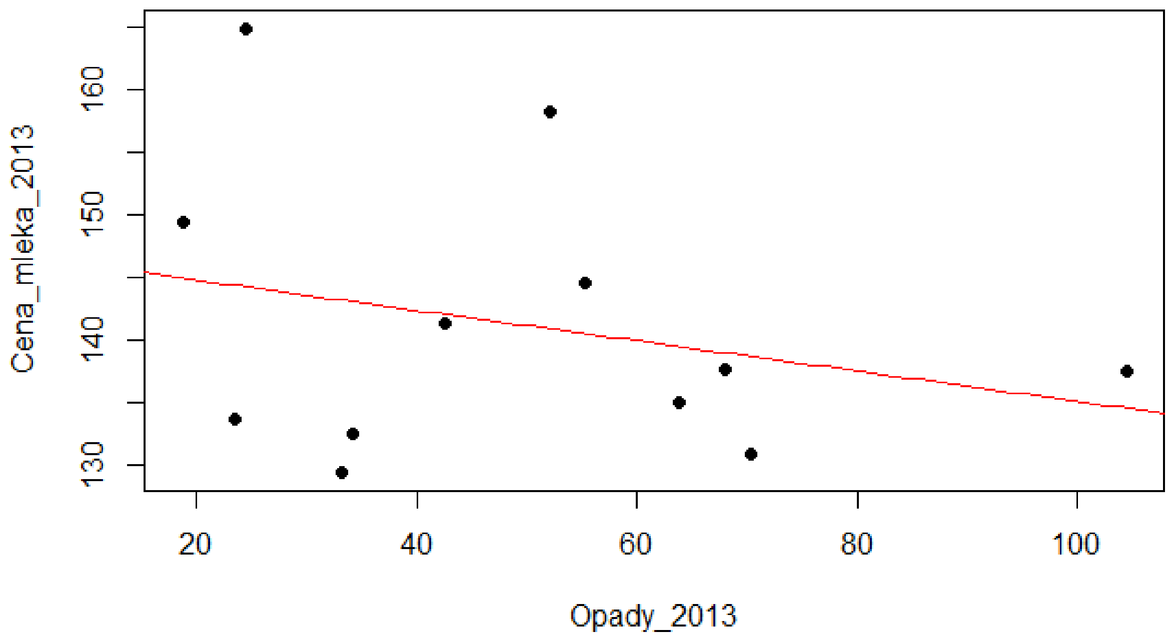
d) 2016 rok



- znikoma korelacja równa 0.0994
- wzór funkcji: ***y = 0.0651*x + 45.6832***
- wpływ opadów na cenę ziemniaków nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena ziemniaków wzrośnie o około 7 gr

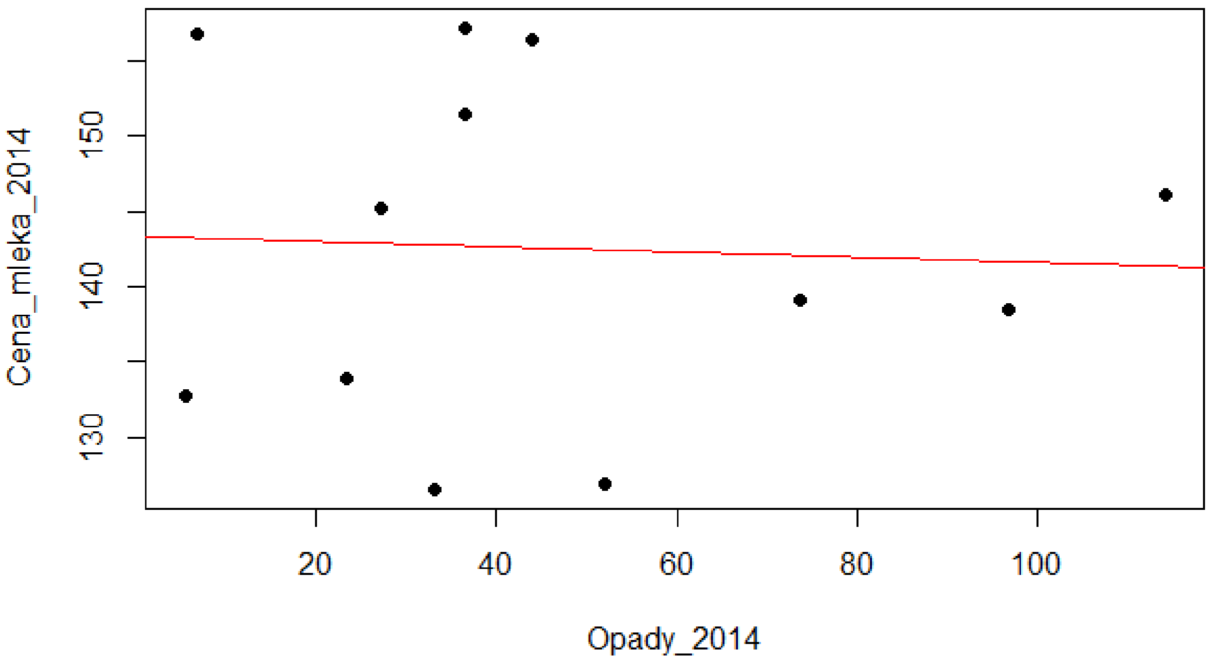
4.4 Zależność ceny mleka od opadów

a) 2013 rok



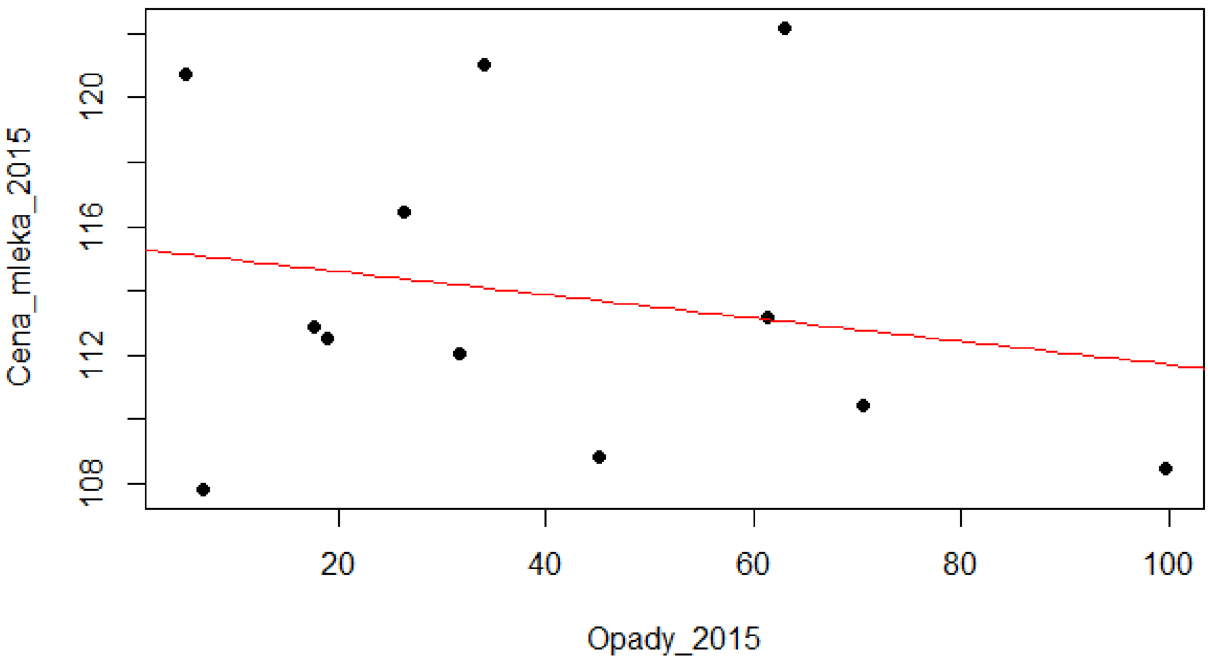
- korelacja słaba równa -0.27
- wzór funkcji: ***y = -0.1212*x + 147.2404***
- wpływ opadów na cenę mleka nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena mleka spadnie o około 12 gr

b) 2014 rok



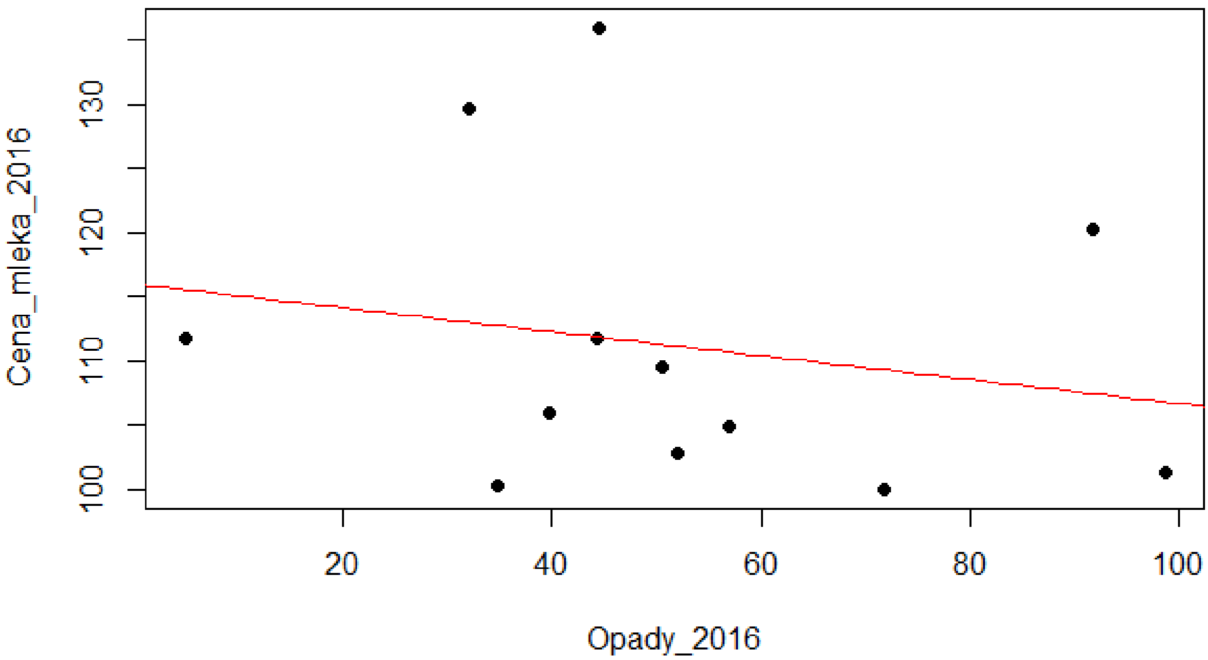
- znikoma korelacja równa -0.0518
- wzór funkcji: ***y = -0.01744*x + 143.36226***
- wpływ opadów na cenę mleka nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena mleka spadnie o około 2 gr

c) 2015 rok



- korelacja słaba równa -0.204
- przy niższej ilości opadów, cena mleka jest wyższa
- wzór funkcji: ***y = -0.0362*x + 115.3211***
- wpływ opadów na cenę mleka nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena mleka spadnie o około 4 gr

d) 2016 rok



- korelacja słaba równa -0.205
- wzór funkcji: ***y = -0.09342*x + 116.01573***
- wpływ opadów na cenę mleka nie jest liniowy, natomiast z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena mleka spadnie o około 9 gr

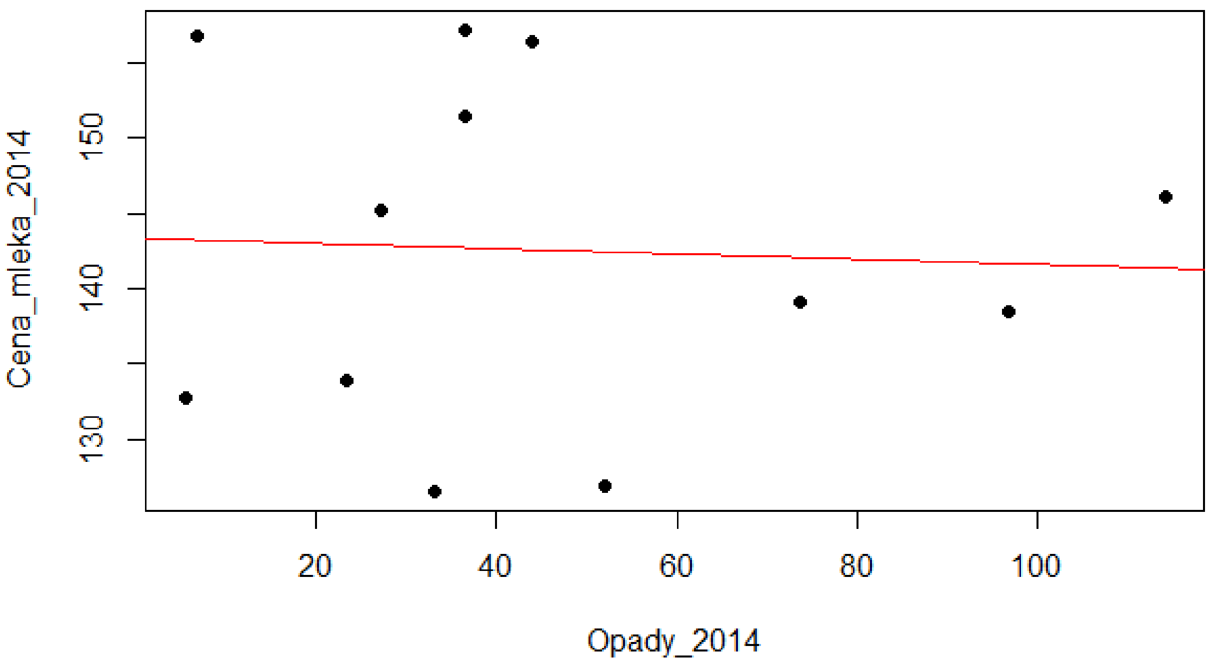
5. Inne funkcje przybliżające układ punktów

W poprzednim rozdziale próbowaliśmy dopasować funkcję liniową w chmurę punktów, stosując regresję liniową, czyli funkcję typu ***y=ax+b***.

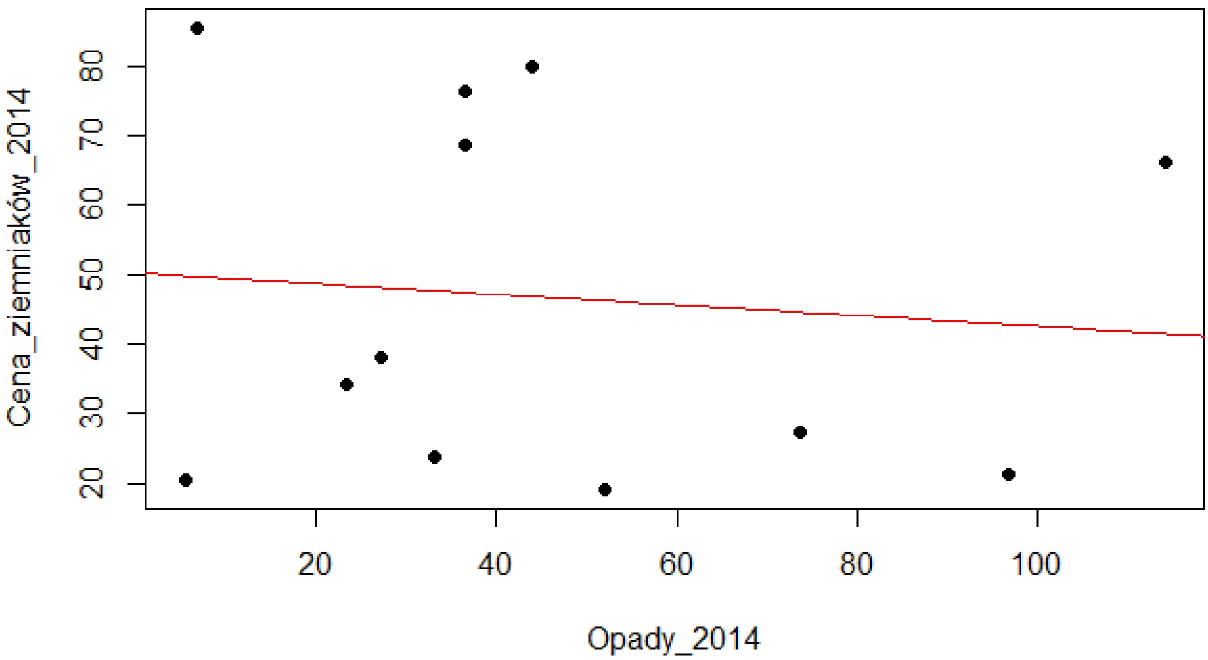
W tym rozdziale spróbujemy na przykładowych wykresach przeskalować dane tak, aby można było w sposób dokładniejszy wpasować tę funkcję. Dzięki temu chcemy otrzymać korelację pomiędzy analizowanymi zmiennymi. Wykorzystamy w tym celu funkcje nieliniowe (np. logarytmiczną, kwadratową itp.), aby w lepszy, niż funkcja liniowa, sposób określić zależność między punktami na wykresie.

Do analizy za pomocą funkcji nieliniowych wybrano zależności:

Cena mleka w 2014 roku zależna od opadów. Korelacja = -0.0518.



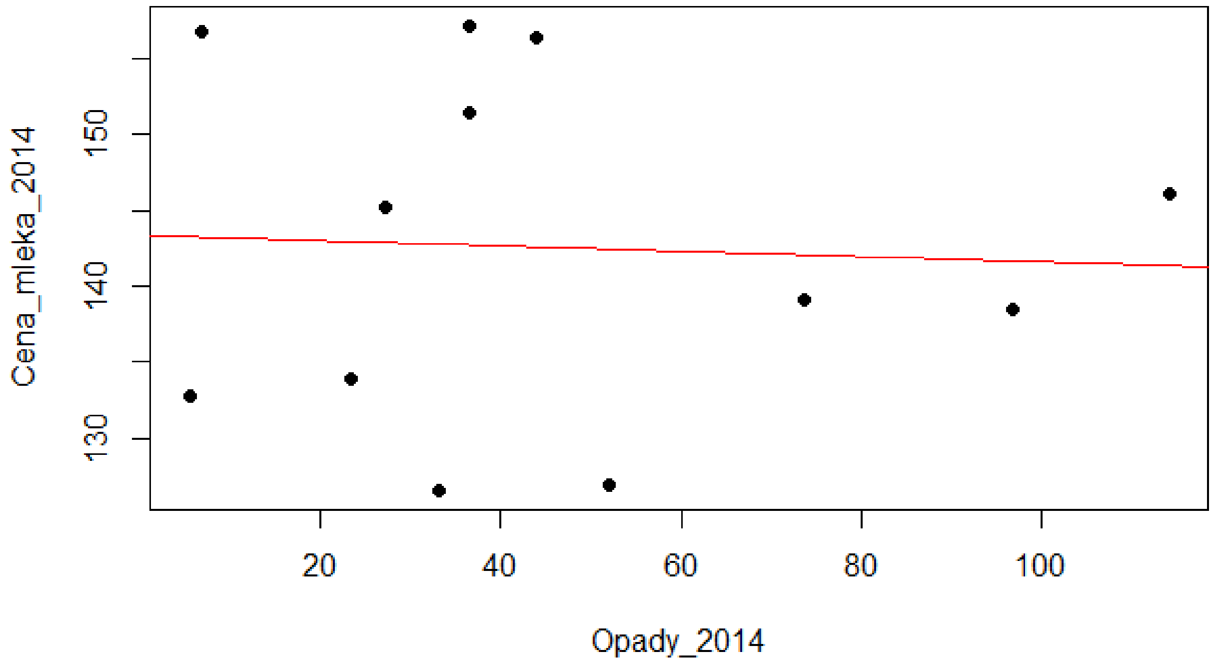
Cena ziemniaków w 2014 roku zależna od opadów. Korelacja = -0.0974



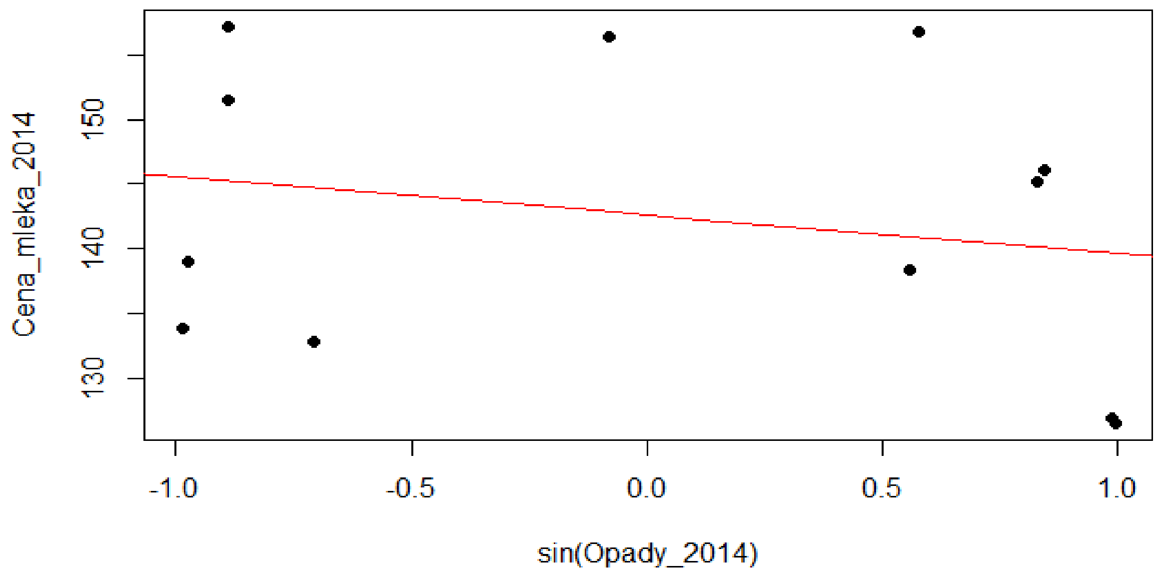
Czy istnieją nieliniowe funkcje, które mogą lepiej wpasować się w chmurę punktów, kiedy korelacja jest bliska zeru? W powyższym przykładzie, obrazującym zależność ceny mleka od opadów, na pierwszy rzut oka ciężko nam będzie stwierdzić czy lepszy będzie wykres paraboli, logarytmiczny czy też sinusoidy.

Metodą prób i błędów będziemy próbowali przekształcić dane tak, aby punkty układały się w sposób umożliwiający dopasowanie takiej funkcji liniowej (bądź nieliniowej), która możliwie najlepiej będzie modelować wszystkie punkty. Spróbujemy uzyskać dzięki temu silniejszą korelację analizowanych zmiennych.

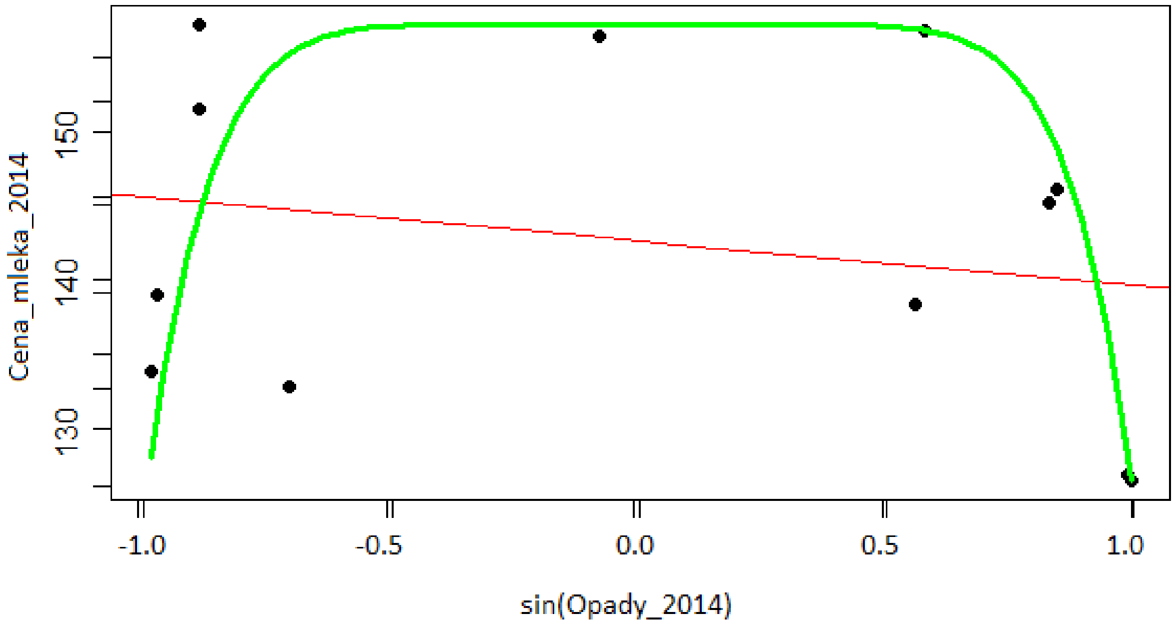
Weźmy pierwszą z zależności z korelacją blisku zera i spróbujmy odpowiednio przekształcić zmienną Opady_2014:



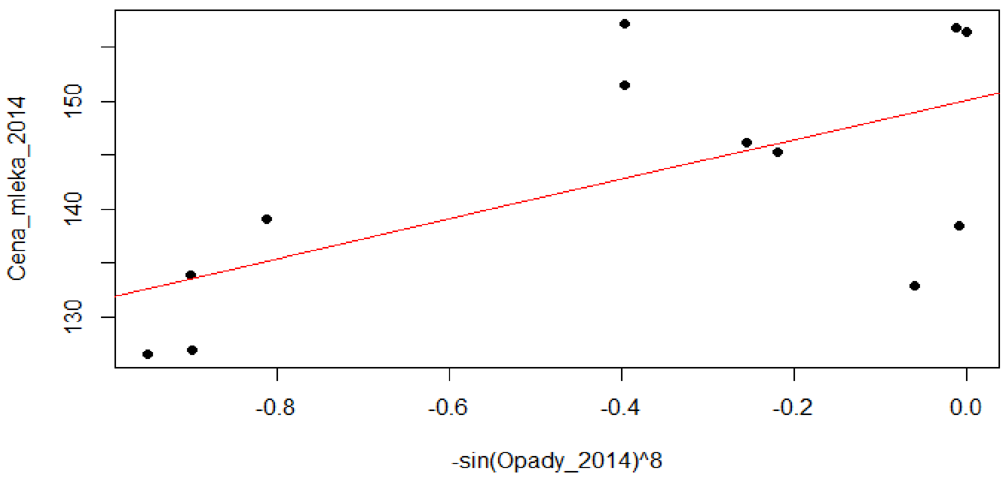
Korelacja wynosi tu -0.0518. Zdecydowanie nie jest to wartość, która nas zadowala. Przekształćmy ten wykres, przeskalowując Opady_2014 funkcją sinus. Mamy wtedy funkcję: ***y = a*sin(x) + b***



Na nowo powstałym wykresie widzimy, że w chmurę punktów najlepiej wpasuje się funkcja paraboli:



Dane z opadami zostały przeskalowane z ***x*** na ***sin(x)***. Funkcja, która najlepiej wpasowała się w punkty na wykresie po tym przeskalowaniu to ***-x^8***. Po kolejnym przeskalowaniu danych, tym razem funkcją ***-x^8*** otrzymano, że cena mleka zależy od opadów wg następującej zależności: ***y = -sin(x)^8 + 155.***
Korelacja dla tej funkcji wynosi -0.6232 (poprzednio -0.0518), co można przedstawić na następującym wykresie:

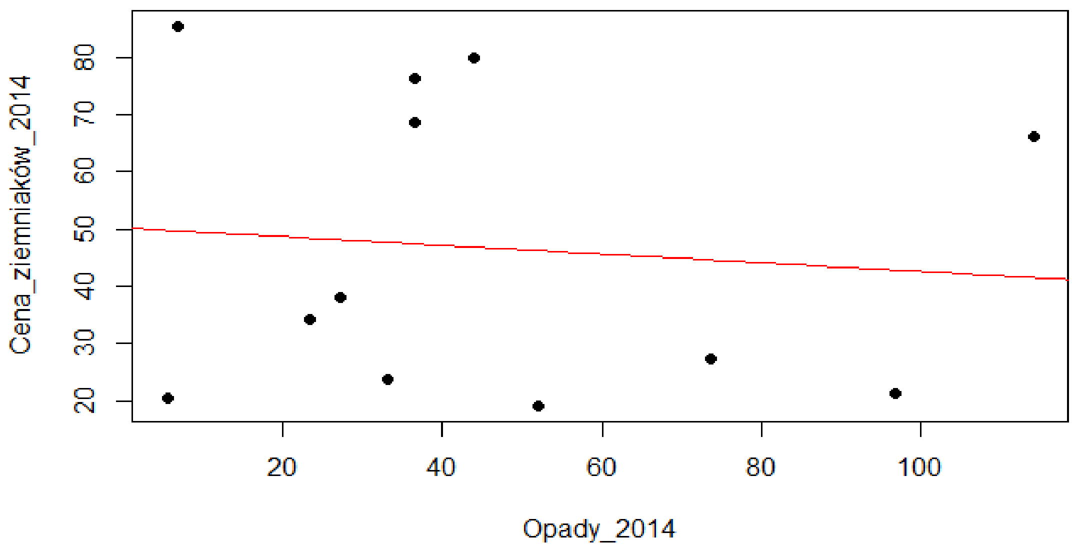


- korelacja będzie wynosiła tyle samo: -0.6232
- zamiast funkcji nieliniowej, otrzymaliśmy liniową: ***y = 18.453* $\bar{x}$ + 150.113***, ale dla przeskalowanej zmiennej
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $4.47*10^{-12}$ przy ***$\bar{x}$*** oraz 0.0304 przy wyrazie wolnym.
  Zakładając poziom istotności alfa = 0.05, zarówno pierwszy jak i drugi współczynnik ma prawdopodobieństwo mniejsze niż alfa
- *R*-squared = 0.3884. Z tego wynika, że w modelu w około 39% przeskalowane opady wyjaśniają cenę mleka
- *p*-value = 0.0304

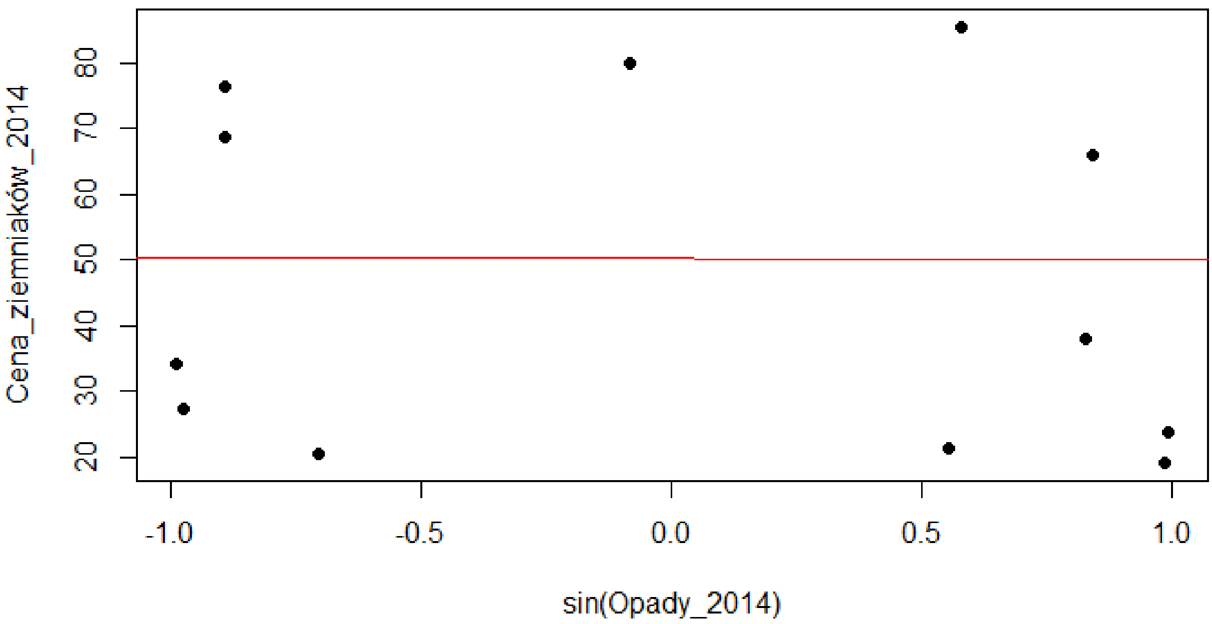
Zatem w tym przykładzie udało się znaleźć funkcję, która pozwoliła nam tak przekształcić dane, aby dopasować funkcję lepiej przybliżającą układ punktów.

Warto również zwrócić uwagę na wyraz wolny w powyższych wzorach. Przy funkcji paraboli „na oko” zauważyliśmy, że najlepiej pasowała liczba ***155***. W powyższym wykresie, gdzie zrobiliśmy przeskalowanie funkcją ***-sin(x)^8***, możemy zauważyć, że przypisana liczba to ***150.113***. Jest to jedna z możliwych funkcji, których wykresem jest parabola. Być może przy funkcji o innej potędze otrzymalibyśmy lepsze dopasowanie.

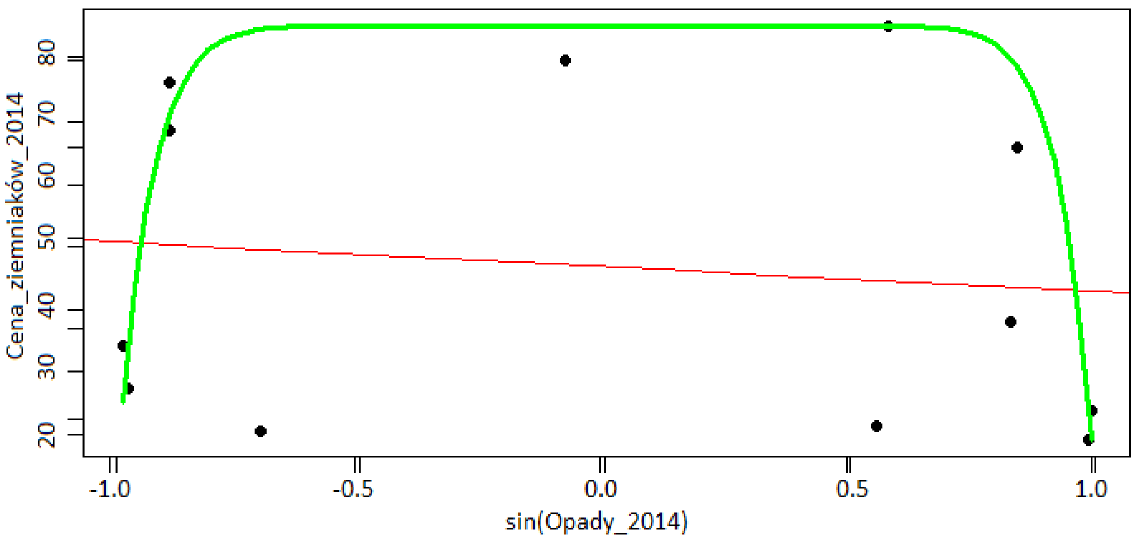
W bardzo podobny sposób będzie wyglądało przekształcenie wykresu zależności ceny ziemniaków od opadów w tym samym roku:



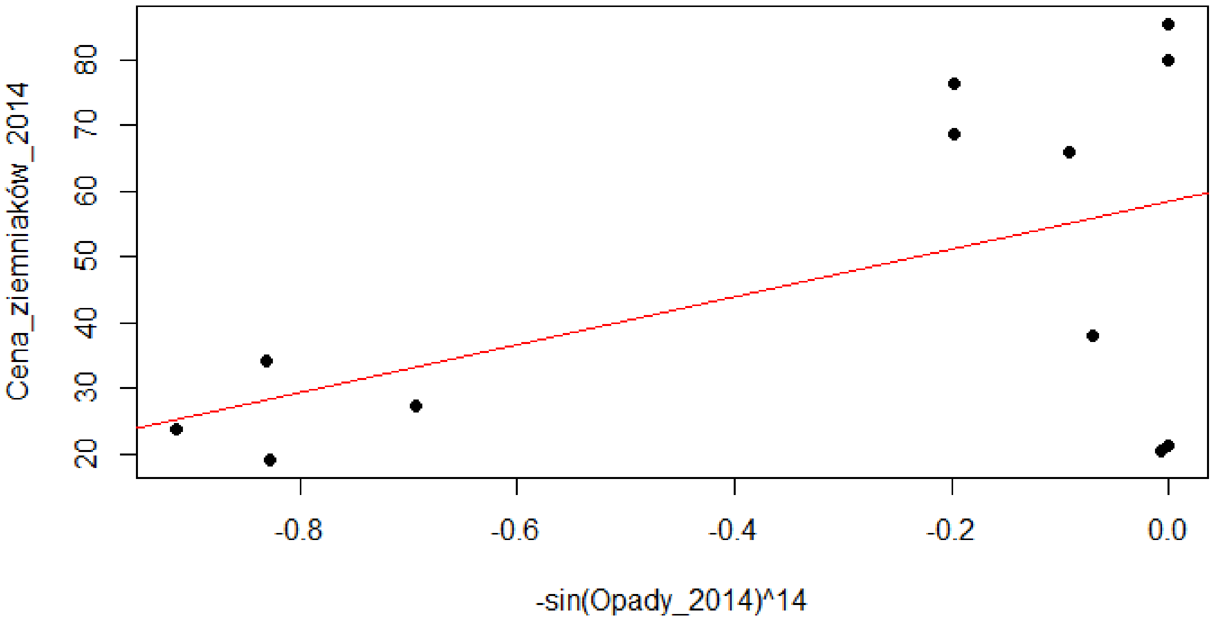
Korelacja wynosi tu -0.0974. Przekształćmy ten wykres, przeskalowując opady funkcją sinus. Mamy wtedy funkcję: ***y = a*sin(x) + b***



Na nowo powstałym wykresie widzimy, że w chmurę punktów najlepiej wpasuje się funkcja paraboli:



Dane z opadami zostały przeskalowane z ***x*** na ***sin(x),*** podobnie jak w poprzednim przykładzie. Funkcja, która najlepiej wpasowała się w punkty na wykresie po tym przeskalowaniu to ***-x^14***. Po kolejnym przeskalowaniu danych, tym razem funkcją ***-x^14*** otrzymano, że cena ziemniaków zależy od opadów wg następującej zależności: ***y = -sin(x)^14 + 85.*** Korelacja dla tej funkcji wynosi 0.5209 (poprzednio -0.0974), co można przedstawić na następującym wykresie:



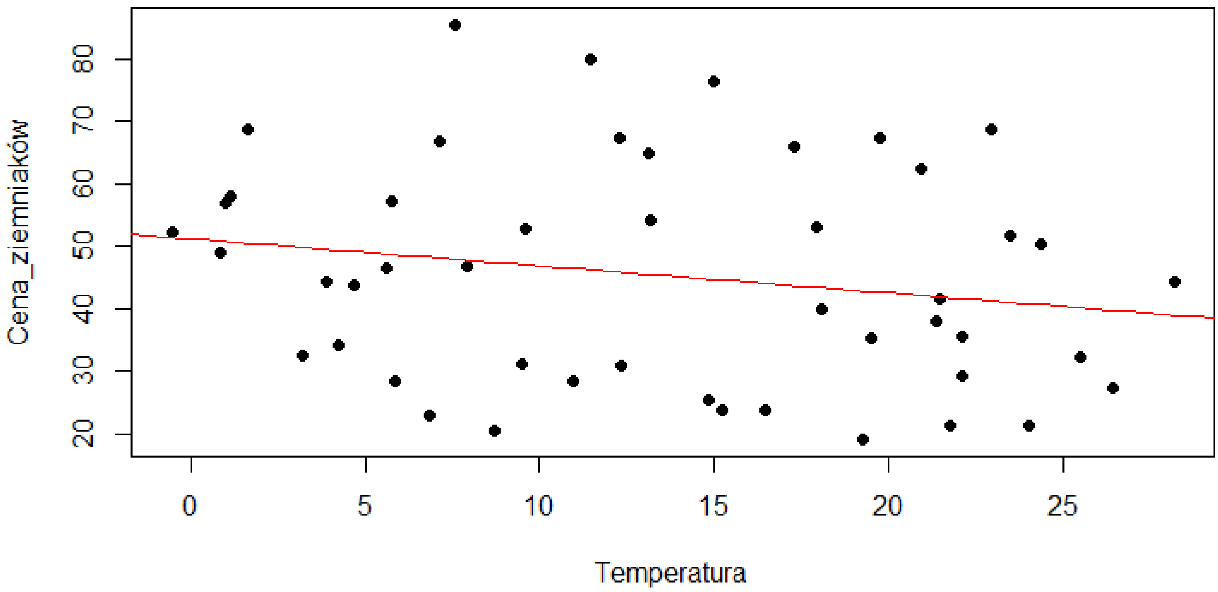
- korelacja będzie wynosiła tyle samo: 0.5209
- zamiast funkcji nieliniowej, otrzymaliśmy liniową: ***y = 36.315** $\bar{x}$ *+ 58.328***, ale dla przeskalowanej zmiennej
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $7.52*10^{-5}$ przy $\bar{x}$ oraz 0.0825 przy wyrazie wolnym.
  Zakładając poziom istotności alfa = 0.05, pierwszy współczynnik ma prawdopodobieństwo błędu mniejsze od alfa, natomiast drugi większe
- *R*-squared = 0.2713. Z tego wynika, że w modelu w około 27% przeskalowane opady wyjaśniają cenę ziemniaków
- *p*-value = 0.08249

6. Badanie zależności w całym okresie 2013-16

W czwartym rozdziale próbowaliśmy zbadać zależności cen produktów od czynników pogody oddzielnie dla poszczególnych lat.
W tym rozdziale zbadamy zależność w całym badanym okresie. Rozważymy 4 przypadki:

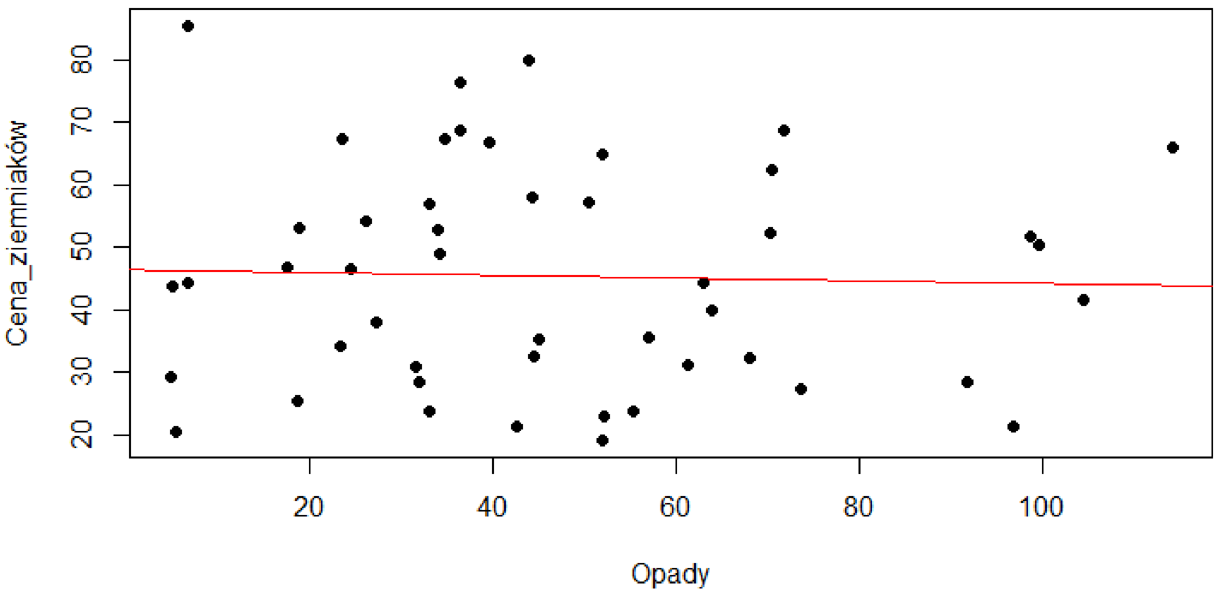
- zależność ceny ziemniaków od temperatury,
- zależność ceny ziemniaków od opadów,
- zależność ceny mleka od temperatury,
- zależność ceny mleka od opadów.

6.1 Zależność ceny ziemniaków od temperatury



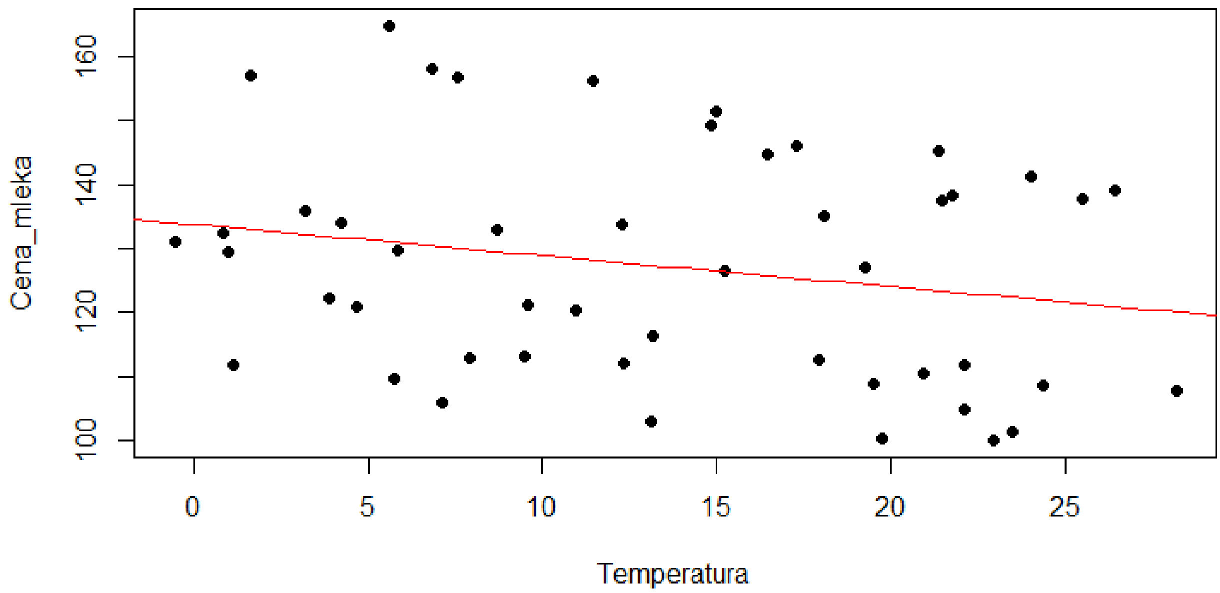
- Korelacja = -0.1977
- wzór funkcji: ***y = -0.4325*x + 51.1789***
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $1.43*10^{-13}$ przy ***x*** oraz 0.178 przy wyrazie wolnym
- z otrzymanego modelu regresji wynika, że wraz ze wzrostem temperatury o 1℃ cena ziemniaków spadnie o około 43 gr

6.2 Zależność ceny ziemniaków od opadów



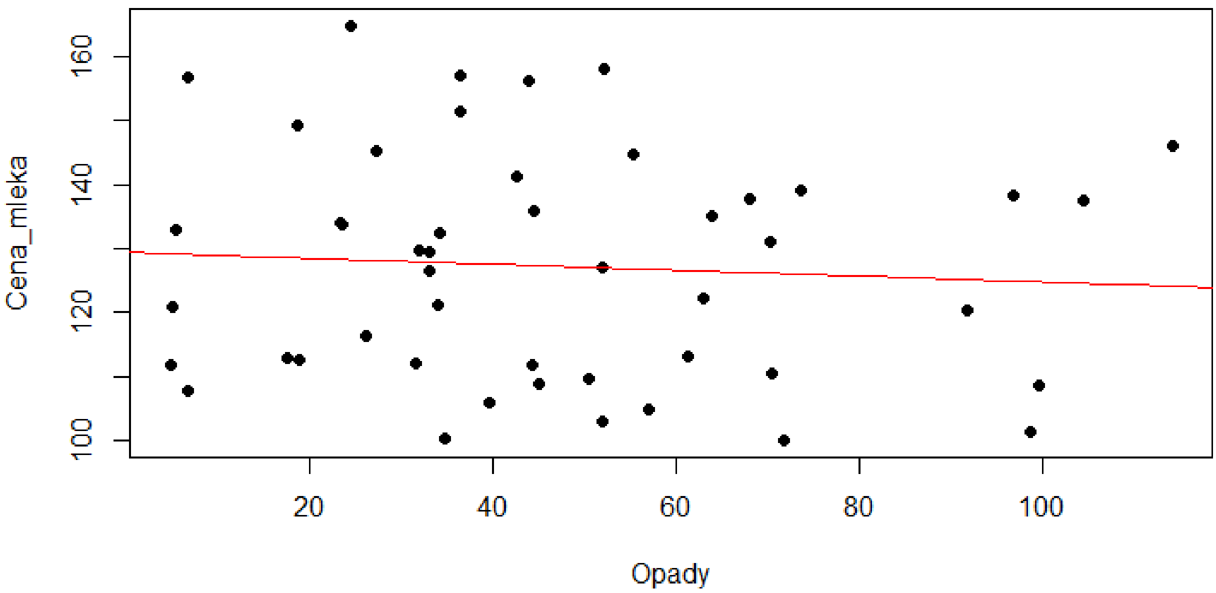
- Korelacja = -0.0349
- Wzór funkcji: ***y = -0.0222*x + 46.3988***
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $7.07 \cdot 10^{-12}$ przy ***x*** oraz 0.814 przy wyrazie wolnym
- z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena ziemniaków spadnie o około 2 gr

6.3 Zależność ceny mleka od temperatury



- Korelacja = -0.2208
- Wzór funkcji: ***y = -0.4863*x + 133.7633***
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $2*10^{-16}$ przy ***x*** oraz 0.132 przy wyrazie wolnym
- z otrzymanego modelu regresji wynika, że wraz ze wzrostem temperatury o 1℃ cena mleka spadnie o około 49 gr

6.4 Zależność ceny mleka od opadów



- Korelacja = -0.0725
- Wzór funkcji: ***y = -0.0464*x + 129.3914***
- prawdopodobieństwo błędu w szacowaniu współczynników w modelu: $2*10^{-16}$ przy ***x*** oraz 0.624 przy wyrazie wolnym
- z otrzymanego modelu regresji wynika, że wraz ze wzrostem opadów o 1 mm cena mleka spadnie o około 5 gr

7. Podsumowanie

Pod uwagę wzięliśmy następujące analizy:

- **zależność ceny ziemniaków od temprartury w 2013 roku,**
- zależność ceny ziemniaków od temprartury w 2014 roku,
- zależność ceny ziemniaków od temprartury w 2015 roku,
- zależność ceny ziemniaków od temprartury w 2016 roku,
- zależność ceny mleka od temperatury w 2013 roku,
- zależność ceny mleka od temperatury w 2014 roku,
- **zależność ceny mleka od temperatury w 2015 roku,**
- **zależność ceny mleka od temperatury w 2016 roku,**
- zależność ceny ziemniaków od opadów w 2013 roku,
- zależność ceny ziemniaków od opadów w 2014 roku,
- zależność ceny ziemniaków od opadów w 2015 roku,
- zależność ceny ziemniaków od opadów w 2016 roku,
- zależność ceny mleka od opadów w 2013 roku,
- zależność ceny mleka od opadów w 2014 roku,
- zależność ceny mleka od opadów w 2015 roku,
- zależność ceny mleka od opadów w 2016 roku.

Na czerwono zostały wyróżnione zależności, gdzie poszczególne czynniki pogodowe statystycznie miały znaczący wpływ na ceny określonych produktów. Jak widać, jest to tylko niewielki procent wszystkich przypadków. W pozostałych sytuacjach wybrane czynniki pogodowe w żaden sposób nie wpływała na kształtowanie się cen. Możemy zatem w sposób ogólny stwierdzić, że każdy z wybranych czynników pogodowych nie ma znaczącego wpływu na ceny produktów. Korelacja dla poszczególnych okresów była słaba lub bardzo słaba.

Kiedy przeprowadzona została analiza dla całego okresu 2013-16, otrzymaliśmy:

- zależność ceny ziemniaków od temprartury, korelacja = -0.1977,
- zależność ceny mleka od temperatury, korelacja = -0.2208,
- zależność ceny ziemniaków od opadów, korelacja = -0.0349,
- zależność ceny mleka od opadów, korelacja = -0.0725.

We wszystkich przypadkach korelacja jest bardzo słaba, jednak gdy spojrzymy na dokładniejsze wartości, zauważymy, że w przypadku zależności ceny mleka od temperatury, gdzie w latach 2015 oraz 2016 mieliśmy dużą zależność, moduł korelacji jest największy. Na drugim miejscu jest z kolei zależność ceny ziemniaków od temperatury, gdzie w 2013 roku ceny były w umiarkowanym stopniu zależne od temperatury.

W celu zwiększenia korelacji, dane zostały przeskalowane na różne sposoby. Jest to pewien sposób, jednak zajęło nam to bardzo dużo czasu. Praktycznie wszystkie wykresy trzeba było „na oko” analizować, pod kątem doboru funkcji, którą należałoby przeskalować dane tak, aby punkty na wykresie układały się wzdłuż linii prostej. Najczęściej były to funkcje nieliniowe.
Wniosek z tego jest taki, że zależność cen od czynników pogodowych lepiej opisują funkcje nieliniowe.

8. Future work

Skoro pogoda nie ma wpływu na ceny produktów, to równie dobrze moglibyśmy mieć przez cały rok 30 stopni i brak opadów, a ceny byłyby takie same. Tak oczywiście nie jest. Na przyszłość warto byłoby dokładniej zagłębić się i przeanalizować, skąd dokładnie importowane są poszczególne produkty. Analiza, jaka została wykonana, była dla Zielonej Góry. Natomiast mleko czy ziemniaki równie dobrze mogły zostać importowane z innej części kraju bądź zza granicy. I w tych miejscach należałoby sprawdzić jaka panowała w badanym czasie pogoda. W sklepach możemy kupić cytryny, jednak są one importowane w zdecydowanej większości z innych krajów, najczęściej z Hiszpanii. Wtedy dla poprawy wyników można znaleźć dane pogodowe dla Hiszpanii oraz ceny cytryny w Polsce.

Innym ciekawym przykładem może być miejscowość Arica, znajdująca się w Ameryce Południowej, a dokładnie w Chile. Jest to miasto znane z „miasta wiecznej wiosny”. Deszcz potrafi nie padać tam całymi latami. Ostatni deszcz jaki tam spadł, był 10 września 2014. Była to w dodatku niegroźna mżawka. Taki klimat powoduje, że nie ma tam żadnej roślinności, upraw ani sadów. Wokół miasta istnieją tylko tereny pustynne. Zatem praktycznie wszystko, co można tam kupić, jest importowane z innej części świata. Jest to kolejny dowód na to, że pogoda w Arica nie ma wpływu na ceny produktów, szczególnie, że przez cały rok temperatura wynosi 20-25 stopni i jest cały czas słonecznie. Od czasu do czasu na niebie pojawią się tylko niegroźne chmury.

Czasami mamy do czynienia z anomaliami pogodowymi, które nie wynikają wprost z wykresów. W 2017 roku po dość ciepłej zimie wiosna przyszła szybko. Jednak kwiecień oraz maj były wyjątkowo zimne. W wielu miejscach spadł nawet śnieg. Najgorsze były jednak przymrozki, które przyszły do Polski w okolicach 9-10 maja. Szczególnie dotkliwy mróz nawiedził wschodnią Polskę, gdzie zanotowano nawet -5℃ Spowodowało to bardzo duże straty w sadach. W niektórych rejonach wymrożonych zostało nawet ponad 80% owoców. Łatwo było to zauważyć na przykład po truskawkach, których w tym roku było zdecydowanie mniej oraz ich cena była wyższa od przeciętnej. W internecie nie brak artykułów o tym. Jeden z nich do poczytania tu: [14].

Przedstawione przykłady mogą posłużyć w późniejszej analizie zależności cen produktów od warunków pogodowych. Modele dzięki nim stworzone będą bardziej rozbudowane i najprawdopodobniej lepiej dopasowane do rzeczywistości.

9. Literatura

[1] http://www.ekologia.pl/pogoda/polska/lubuskie/zielona-gora/archiwum,zakres,01-01-2016_31-01-2016

[2] https://dane.imgw.pl/

[3] Dane dotyczące cen produktów - http://stat.gov.pl/

[4] https://www.wunderground.com/history/airport/2017/07/26/DailyHistory.html?req_city=Luqa&req_statename=Malta&reqdb.zip=00000&reqdb.magic=272&reqdb.wmo=16597&MR=1

[5] https://www.agrofakt.pl/rolnictwo-a-pogoda-plony-dalej/

[6] https://pl.wikipedia.org/wiki/Wsp%C3%B3%C5%82czynnik_korelacji_Pearsona

[7] http://www.farmer.pl/produkcja-zwierzeca/bydlo-i-mleko/czy-ceny-mleka-zaczna-spadac,72778.html

[8] https://pl.wikipedia.org/wiki/Regresja_(statystyka)

[9] http://www.naukowiec.org/wiedza/statystyka/regresja-liniowa_765.html

[10] https://pl.wikipedia.org/wiki/Tendencja_rozwojowa

[11] http://analizystatystyczne.com/tag/korelacja/

[12] http://www.naukowiec.org/wiedza/statystyka/wspolczynnik-determinacji_736.html

[13] http://www.naukowiec.org/wiedza/statystyka/metoda-najmniejszych-kwadratow_733.html

[14] http://www.dziennikzachodni.pl/wiadomosci/slask/a/kwietniowe-i-majowe-mrozy-zniszczyly-sady-i-uprawy-polskie-owoce-beda-drogie-i-zbierzemy-ich-mniej,12071953/