**Analiza danych parametrów win z jednego rejonu z podziałem na odmiany winogron**

Plik wine\_data.csv zawiera wyniki pomiarów kilku wybranych parametrów win pochodzących z tego samego rejonu Włoch, ale wyprodukowanych z trzech różnych odmian (ang. cultivar) winogron:

* Alcohol (Alkohol) - procentowa zawartości alkoholu w winie.
* Malic acid (kwas jabłkowy) - rodzaj kwasu obecnego w winogronach i innych owocach.
* Ash (popiół) - reprezentuje pozostałości po fermentacji winogron, obejmujące składniki mineralne, takie jak potas i wapń.
* Flavanoids (flawonoidy) - rodzaj związku fenolowego.
* Color intensity (intensywność koloru) - mierzy jak głęboki lub intensywny jest kolor.
* OD280/OD315 (absorbancja przy 280/315 nm wina rozcieńczonego) - miara absorpcji światła przy określonych długościach fal.

Tabela zawiera pomiary przeprowadzone dla 178 win.

**Charakterystyka poszczególnych parametrów.**

Tabela podstawowych statystyk opisowych poszczególnych parametrów:

**A table of numbers and symbols

Description automatically generated with medium confidence**

1. **Alkohol**

Zawartość alkoholu w winach oscyluje w granicach od 11 - 14,8 procent, a średnio ta wartość wynosi 13 procent. Odchylenie standardowe wynosi około 0.8 punktu procentowego, co sugeruje umiarkowane rozproszenie wokół średniej wartości. Skośność danych jest bliska zeru (około -0.05), co wskazuje na minimalną asymetrię w rozkładzie. Kurtoza wynosząca -0.86 sugeruje, że rozkład alkoholu jest nieco bardziej płaski w porównaniu z rozkładem normalnym. A graph of blue dots

Description automatically generatedA chart with different colored squares

Description automatically generated

Analizując wykresy boxplotu, możemy zaobserwować, że dla pierwszej i trzeciej odmiany winogron (Cultivar) średnia zawartość alkoholu jest zbliżona. Odmiana pierwsza ma go minimalnie więcej. Jednakże, dla drugiej odmiany jest ona niższa, choć występują tam kilka wartości odstających, które są prawie tak duże jak w pozostałych przypadkach.

Współczynnik korelacji liniowej wynoszący -0.33 sugeruje istnienie umiarkowanej negatywnej zależności między odmianami winogron a zawartością alkoholu.

1. **Kwas jabłkowy**

Dane dotyczące kwasu jabłkowego wykazują średnią wartość wynoszącą około 2.34, przy odchyleniu standardowym wynoszącym 1.12, co sugeruje umiarkowane rozproszenie wokół średniej. Wartość minimalna kwasu jabłkowego wynosi 0.74, a maksymalna 5.8, co wskazuje na zróżnicowanie w zawartości tego kwasu w badanych winach. Skośność danych wynosząca około 1.03 sugeruje asymetrię w prawo, co może oznaczać większe występowanie wyższych wartości kwasu jabłkowego. Kurtoza o wartości 0.26 wskazuje na lekko spłaszczony rozkład w porównaniu z rozkładem normalnym.

**A graph of a graph showing a line of blue dots

Description automatically generated with medium confidence** A chart with different colored squares and numbers

Description automatically generated

Rozkład kwasu jabłkowego w badanych winach jest zróżnicowany, ale głównie skupiony w niższych wartościach, co sugeruje dominację niższych poziomów kwasowości. Jednak obserwuje się także wartości odstające zarówno w górę, jak i w dół, co wskazuje na pewną złożoność tego parametru. Szczególnie warto zauważyć, że odmiana 3 charakteryzuje się wyższymi wartościami kwasu jabłkowego w porównaniu z odmianami 1 i 2, co może sugerować różnice w procesach fermentacji lub warunkach hodowlanych.

Wartość współczynnika korelacji liniowej wynosząca 0.44 wskazuje na umiarkowany pozytywny związek między zmiennymi.

1. **Popiół**

Popiół w badanych winach wykazuje średnią wartość wynoszącą około 2.37, przy niewielkim rozproszeniu wokół tej średniej (odchylenie standardowe 0.27). Zakres danych oscyluje od 1.36 do 3.23, z wartościami skupionymi przede wszystkim wokół wartości średniej. Skośność danych jest minimalna (-0.18), co sugeruje niewielką asymetrię w rozkładzie. Natomiast wartość kurtozy wynosząca 1.08 wskazuje na nieco bardziej stromy rozkład w porównaniu z rozkładem normalnym, co może oznaczać obecność kilku wartości odstających.

A chart with a number of colored squares

Description automatically generated with medium confidenceA graph with blue dots

Description automatically generated

W analizie popiołu w badanych winach można zauważyć, że średnia wartość dla wszystkich odmian jest zbliżona do ogólnej średniej, co sugeruje pewną spójność w ich zawartościach. Niemniej jednak, odmiana druga wyróżnia się wyraźnie z większymi wartościami odstającymi, przy jednoczesnym utrzymaniu się nieco poniżej średniej ogólnej.

Wartość współczynnika korelacji liniowej wynosząca -0.05 wskazuje na bardzo słaby lub brak związek liniowy między zmiennymi.

1. **Flawonoidy**

Flawonoidy w badanych winach charakteryzują się średnią wartością około 2.03, przy umiarkowanym rozproszeniu wokół tej średniej (odchylenie standardowe 0.99). Wartości wahają się od 0.34 do 5.08, co pokazuje znaczną zmienność w zawartości tego związku. Mimo niewielkiej asymetrii w rozkładzie danych (skośność 0.03), ogólny kształt rozkładu jest stosunkowo zbliżony do rozkładu normalnego, chociaż lekko spłaszczony (kurtoza -0.89).

A diagram of a group of people

Description automatically generated with medium confidenceA diagram of a plant

Description automatically generated

Analiza sugeruje, że zawartość flawonoidów w winach wykazuje wyraźne różnice między odmianami. Odmiana 1 posiada najwyższe wartości, a odmiana 3 - najniższe. Jest zauważalna tendencja liniowa, gdzie wartości zmiennej zależnej maleją wraz ze wzrostem numeru odmiany, co odpowiada relacji ax + b, gdzie a < 0. Współczynnik korelacji liniowej wynoszący -0.85 potwierdza istnienie silnego ujemnego związku między numerem odmiany a zawartością flawonoidów, co oznacza, że im niższy numer odmiany, tym wyższa zawartość flawonoidów.

1. **Intensywność koloru**

Intensywność koloru w badanych winach jest średnio umiarkowana (średnia: 5.06) z dość znacznym rozproszeniem wokół tej wartości (odchylenie standardowe: 2.32). Widoczna jest duża zmienność w intensywności koloru między różnymi winami, od wartości minimalnej 1.28 do maksymalnej 13.00. Skośność dodatnia (0.86) sugeruje pewną asymetrię w prawo w rozkładzie danych. Kurtoza (0.34) wskazuje na to, że rozkład jest lekko bardziej skupiony niż rozkład normalny.

A diagram of a group of colored squares

Description automatically generated with medium confidenceA graph of blue dots

Description automatically generated

Rozkład intensywności koloru w badanych winach wykazuje największą zmienność w porównaniu z innymi zmiennymi, chociaż większość wartości skupiona jest w zakresie od 2 do 9. Odmiana 2 charakteryzuje się najniższymi wartościami w porównaniu z innymi odmianami.

1. **Absorbancja przy 280/315 nm wina rozcieńczonego**

Absorbancja przy 280/315 nm wina rozcieńczonego wykazuje umiarkowany poziom, średnio wynoszący około 2.61, z niewielkim rozproszeniem wokół tej wartości. Badane próbki mają różny stopień absorbancji, mieszcząc się w zakresie od 1.27 do 4.00. Skośność jest lekko ujemna, co wskazuje na lekką asymetrię w lewo w rozkładzie danych, a kurtoza sugeruje, że rozkład jest bardziej płaski niż rozkład normalny.

**A chart with a number of colored squares

Description automatically generated with medium confidenceA graph of blue dots

Description automatically generated**

Analiza danych wskazuje, że wraz z każdym kolejnym typem winogrona absorbancja przy 280/315 nm maleje. Dodatkowo, współczynnik korelacji liniowej wynoszący -0.79 potwierdza silną ujemną zależność między typem winogrona a absorbancją, co oznacza, że im wyższy numer typu winogrona, tym niższa absorbancja.

**Rozróżnienie win wykonanych z poszczególnych odmian winogron**

Parametry Flavanoids i Color intensity wydają się być kluczowe do rozróżnienia odmian, ponieważ występują one w miare na odróżnialnych pułapach w zależności od odmiany. Żeby łatwo odróżnić odmianę 1 i 3 należy porównywać zestawienie parametrów odmiany względem Flavanoids lub OD280/OD315.

Żeby rozróżnić odmianę 2 i 3 należy porównywać zestawienie parametrów odmiany względem Color intensity OD280/OD315. Natomiast żeby rozróżnić odmianę 1 i 2 należy porównywać zestawienie parametrów odmiany względem Alcohol.

**Związki między poszczególnym parametrami win**

A group of dots with black text

Description automatically generatedA graph of a graph showing a number of dots

Description automatically generated with medium confidenceW całym zestawie danych obserwujemy wyraźną korelację oraz wyraźnie nachyloną prostą regresji między zmiennymi OD280/OD315 a Flavanoids. Jednak po podziale danych ze względu na rodzaj winogron użyte do produkcji wina, tylko druga grupa winogron utrzymuje tę zależność.A graph of a line with blue dots

Description automatically generated

W całym zestawie danych obserwujemy również wyraźną korelację oraz wyraźnie nachyloną prostą regresji między zmiennymi Alkohol a Color Intensity. Po podziale danych ze względu na rodzaj winogron użyte do produkcji wina, proste wszystkich grup zachowują kierunek pierwotnej prostej.

A group of dots with black text

Description automatically generatedA graph showing a number of dots

Description automatically generated with medium confidenceA graph of a line with blue dots

Description automatically generated with medium confidenceA group of dots with black text

Description automatically generated