1. Statystyka

1.1 Rola i znaczenie statystyki w analizie danych w łańcuchach dostaw

Statystyka odgrywa kluczową rolę w nowoczesnych łańcuchach dostaw, gdzie efektywne zarządzanie, planowanie i kontrola są niezbędne. Metody statystyczne są wykorzystywane do gromadzenia, analizowania i interpretowania danych, umożliwiając firmom lepsze zrozumienie i optymalizację ich łańcuchów dostaw.

Przedstawmy niektóre z ważnych ról statystyki w analizie łańcucha dostaw.

Statystyki opisowe są kluczem do opisania podstawowych cech serii danych łańcucha dostaw, takich jak średnia, odchylenie standardowe, mediana, kwartyle i inne miary. Narzędzia te pomagają nam zrozumieć rozkład i charakterystykę danych, takich jak średni czas dostawy, ilości w magazynie i średnie koszty, co przyczynia się do lepszego zrozumienia i zarządzania łańcuchem dostaw.

Ponadto techniki statystyczne, takie jak regresja, analiza szeregów czasowych i wzorce analityczne danych, są wykorzystywane do przewidywania przyszłych zdarzeń i trendów w łańcuchach dostaw. Obejmuje to prognozowanie popytu, zapasów i czasu dostawy, umożliwiając lepsze planowanie i dostosowanie dostaw.

Statystyki odgrywają kluczową rolę w identyfikowaniu wzorców w danych, umożliwiając lepsze zrozumienie zachowania łańcucha dostaw, w tym wzorców sezonowych, trendów i cykli popytu.

Optymalizacja zapasów to kolejny kluczowy obszar, w którym statystyki pomagają określić optymalne ilości zamówień, które minimalizują koszty magazynowania i zamawiania, przy użyciu metod takich jak EOQ (Economic Order Quantity).

Ponadto, statystyki są również wykorzystywane do oceny ryzyka związanego z łańcuchem dostaw, takiego jak prawdopodobieństwo opóźnień w dostawach, uszkodzeń podczas transportu i innych potencjalnych problemów.

Dzięki monitorowaniu statystycznemu i kontroli procesów identyfikujemy odchylenia od standardów, co pozwala nam poprawić jakość i wydajność procesów łańcucha dostaw.

Ponadto statystyki są wykorzystywane do monitorowania i poprawy jakości produktów i usług w łańcuchu dostaw, w tym kontroli jakości u dostawców.

Wreszcie, statystyki są kluczowym narzędziem do podejmowania bardziej świadomych decyzji dotyczących zaopatrzenia, zapasów, wyboru dostawców i innych aspektów zarządzania dostawami, przyczyniając się do wydajnego i skutecznego działania całego łańcucha dostaw.

W analizie łańcucha dostaw statystyki są wykorzystywane do optymalizacji procesów, redukcji kosztów, zwiększenia wydajności i poprawy zadowolenia klientów. Umożliwia lepsze zrozumienie dynamiki łańcucha dostaw i lepsze zarządzanie ryzykiem, co ma kluczowe znaczenie dla pomyślnego funkcjonowania firm i organizacji w dzisiejszym globalnym środowisku.

1.2 Podstawowe pojęcia statystyki

Zmienne

Zmienne są podstawowymi elementami składowymi w statystyce, ponieważ reprezentują właściwości lub cechy, które są mierzone lub obserwowane w badaniu, eksperymencie lub próbie danych. Zmienne są niezbędne do zrozumienia i analizy danych, ponieważ pozwalają badaczom, analitykom i statystykom opisywać, analizować i rozumieć zjawiska.

Ważne jest, aby zrozumieć różne rodzaje zmiennych i ich znaczenie w statystyce.

Zmienne jakościowe (nominalna, kategorialna/porządkowa) to zmienne reprezentujące cechy jakościowe, czyli niemierzalne na skali liczbowej lub zmienne, które klasyfikują rozłącznie jednostki danych. Przykłady obejmują płeć (mężczyzna, kobieta), kolor oczu (niebieski, brązowy, zielony) lub typ samochodu (sedan, kombi, SUV). Zmienne jakościowe są często przydatne do opisywania cech demograficznych lub innych cech.

Zmienne ilościowe (liczbowe) to zmienne reprezentujące wartości liczbowe, które można policzyć lub zmierzyć i które można posortować w pewnym porządku matematycznym. Przykłady obejmują wiek w latach, wzrost, temperaturę, dochód lub wyniki ankiet. Zmienne ilościowe są często wykorzystywane do analizy i ilościowego badania zjawisk.

Zmienne zależne i niezależne. Zmienna zależna to ta, którą chcemy zbadać, zmierzyć lub przewidzieć, podczas gdy zmienna niezależna to ta, która ma wpływać na zmienną zależną. Na przykład, jeśli chcemy zbadać, czy poziom wykształcenia wpływa na dochód, dochód jest zmienną zależną, a poziom wykształcenia zmienną niezależną.

Zmienne dyskretne i ciągłe. Zmienne można również podzielić na dyskretne i ciągłe. Zmienne dyskretne mają ograniczony zestaw możliwych wartości i są zwykle reprezentowane przez liczby całkowite. Przykładem może być liczba dzieci w rodzinie, gdzie możliwe wartości to 0, 1, 2 itd. Z drugiej strony zmienne ciągłe mają nieskończoną liczbę możliwych wartości i są zwykle mierzone za pomocą liczb dziesiętnych. Przykładem jest wzrost osób, gdzie możliwa jest nieskończona liczba wartości w danym zakresie.

Zmienne są podstawowymi narzędziami badań i analizy danych. Zrozumienie i prawidłowe zdefiniowanie zmiennych ma kluczowe znaczenie dla przeprowadzania analiz statystycznych i badania zjawisk w badaniach. Zmienne pozwalają badaczom wyrażać i kwantyfikować różne aspekty rzeczywistości, umożliwiając lepsze zrozumienie zjawisk, podejmowanie decyzji i przewidywanie przyszłych zdarzeń. Pozwalają również na wykorzystanie różnych technik statystycznych do testowania hipotez, tworzenia prognoz i lepszego zrozumienia związków przyczynowych między zmiennymi.

1.3 Podstawowe pojęcia statystyczne z przykładami

Średnia arytmetyczna (średnia)

Średnia, znana również jako **średnia arytmetyczna**, jest jedną z podstawowych miar statystycznych. Średnia jest średnią arytmetyczną wszystkich wartości w zestawie danych. Oblicza się ją poprzez zsumowanie wszystkich danych, a następnie podzielenie przez liczbę danych.

Obliczanie średniej arytmetycznej:

- Zsumowanie wszystkich wartości w zbiorze danych.
- Podzielenie sumy przez liczbę danych w zbiorze.
- Równanie do obliczania średniej $(\bar{x})$ to: $\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_n) / n$

Gdzie: $\bar{x}$ jest średnią arytmetyczną. $x_1, x_2, x_3, \ldots, x_n$ są wartościami w zbiorze danych. n to liczba wartości w zbiorze danych.

Przykład:

Wyobraźmy sobie zbiór danych reprezentujący oceny uczniów z egzaminu z matematyki: 80, 85, 90, 75, 95. Aby obliczyć średnią, należy dodać wszystkie te wartości i podzielić przez liczbę ocen, która w tym przypadku wynosi 5:

Średnia arytmetyczna = (80 + 85 + 90 + 75 + 95) / 5 = 425 / 5 = 85

Tak więc średni wynik ucznia wynosi 85. Średnia jest przydatna do pomiaru tendencji centralnej danych i daje nam przybliżony obraz tego, czego możemy się spodziewać jako „typowej" wartości w zestawie danych. Jednak średnia może się znacznie zmienić, jeśli w danych występują wartości odstające. Dlatego ważne jest, aby znać inne miary statystyczne, takie jak mediana i moda, aby lepiej zrozumieć rozkład danych.

Mediana

Mediana to pojęcie statystyczne używane do pomiaru środkowej wartości cechy. W przypadku parzystej liczby danych o różnych wielkościach liczbowych połowa danych ma wartości mniejsze lub równe medianie, a druga połowa ma wartości większe lub równe medianie. Mediana jest jedną z podstawowych miar tendencji centralnej w statystyce i służy do opisu rozkładu danych, zwłaszcza gdy dane są skośne lub zawierają wartości odstające

Jak obliczyć medianę?

- Najpierw należy posortować zestaw danych od najmniejszej do największej wartości.
- Jeśli liczba danych jest parzysta (n), to mediana jest średnią dwóch środkowych wartości. Oznacza to, że mediana jest równa średniej wartości na pozycji $n/2$ i $(n/2 + 1)$ gdy dane są posortowane w porządku rosnącym.
- Jeśli liczba danych jest nieparzysta, wartość mediany znajduje się na środkowej pozycji.

Przykład:

Wyobraźmy sobie następujący zestaw danych przedstawiający liczbę godzin snu w danym okresie: 7, 6, 5, 8, 6, 9, 7

Najpierw uporządkujmy dane w kolejności rosnącej: 5, 6, 6, 7, 7, 8, 9

Ponieważ liczba danych jest nieparzysta (7), medianą będzie wartość na środkowej pozycji, która jest czwartą wartością w uporządkowanym zestawie danych. Zatem mediana w tym

przypadku wynosi 7 godzin. Oznacza to, że połowa osób w tym zbiorze danych śpi 7 lub mniej godzin, podczas gdy druga połowa śpi 7 lub więcej godzin.

Moda (dominanta)

Moda (dominanta) jest jednym z podstawowych wskaźników statystycznych używanych do pomiaru tendencji centralnej zbioru danych. Moda (dominanta) reprezentuje wartość, która występuje najczęściej w zbiorze danych. Jest to wartość, która ma najwyższą częstotliwość występowania spośród wszystkich wartości w zbiorze danych.

Moda (dominanta) jest przydatna do identyfikowania najczęstszych wartości w zbiorze danych i jest szczególnie przydatna podczas analizowania zmiennych jakościowych (kategorialnych), w których wartości są nienumeryczne.

Jeśli w zestawie danych występuje wiele mód (wiele wartości występujących z podobną maksymalną częstotliwością), mówimy o rozkładzie wielomodalnym. Jeśli wszystkie dane mają taką samą częstotliwość występowania, zestaw danych nie ma mody.

Przykład: wyobraźmy sobie zbiór danych reprezentujący kolory samochodów na parkingu:

Czerwony, Niebieski, Czerwony, Zielony, Niebieski, Niebieski, Czerwony

W tym przypadku modą jest „Czerwony", ponieważ ta wartość występuje najczęściej (trzy razy), podczas gdy „Niebieski" i „Zielony" występują rzadziej.

Moda (dominanta) jest prosta do obliczenia, ponieważ po prostu identyfikuje wartość o najwyższej częstotliwości występowania w zbiorze danych. Moda jest używana do opisywania charakterystycznych wartości w danych i może być przydatna w zrozumieniu, która wartość jest najbardziej charakterystyczna dla określonej sytuacji lub grupy.

Zakres zmienności (VR, zakres, interwał, amplituda)

Różnica między wartościami maksymalnymi i minimalnymi w zestawie danych jest pojęciem statystycznym zwanym zakresem. Mierzy on, jak duża jest różnica między wartościami maksymalnymi (maksimum) i minimalnymi (minimum) w zestawie danych. Zakres to prosty sposób na oszacowanie zakresu wartości w zestawie danych i zmierzenie zmienności między wartościami minimalnymi i maksymalnymi.

Obliczenie marginesu zmienności jest proste:

- Najpierw należy znaleźć wartość minimalną (min) i maksymalną (max) w zbiorze danych.
- Następnie obliczyć różnicę między wartością maksymalną i minimalną (max - min).

Przykład: wyobraźmy sobie zbiór danych reprezentujący wiek uczestników wydarzenia: 20, 25, 30, 35, 40. Aby obliczyć margines zmienności, należy najpierw znaleźć wartość minimalną (20) i maksymalną (40) w zbiorze danych. Następnie obliczamy różnicę między wartością maksymalną i minimalną: VR = 40 - 20 = 20

Zatem margines zmienności w tym przypadku wynosi 20 lat. Oznacza to, że różnica między najstarszym a najmłodszym uczestnikiem wynosi 20 lat.

Rozkład wariancji jest przydatny do oszacowania zakresu wartości w zbiorze danych, ale jest dość prosty i nie uwzględnia wszystkich wartości w zbiorze danych. Do bardziej szczegółowej analizy zmienności i rozproszenia danych powszechnie stosuje się inne miary statystyczne, takie jak wariancja lub kwartyle.

Wariancja i odchylenie standardowe

Wariancja to średnia kwadratów odchyleń od średniej. Jest to kwadrat odchylenia standardowego. **Odchylenie standardowe** jest miarą statystyczną używaną do pomiaru rozproszenia lub zmienności w zestawie danych. Informuje, jak daleko wartości są od średniej (przeciętnej) w zestawie. Odchylenie standardowe jest jedną z najczęściej używanych miar rozproszenia w statystyce i jest obliczane poprzez obliczenie pierwiastka kwadratowego z wariancji (wariancji).

Obliczanie odchylenia standardowego:

- Najpierw należy obliczyć zmienność (wariancję). Zmienność (wariancję) oblicza się, biorąc średnią z kwadratów różnic od średniej.
- Po uzyskaniu wartości wariancji (σ^2), pierwiastkując ją, należy obliczyć odchylenie standardowe. Odbywa się to poprzez wyznaczenie pierwiastka kwadratowego z σ^2:

Odchylenie standardowe $\sigma = \sqrt{\sigma^2}$.

Odchylenie standardowe mierzy, jak rozproszone są wartości wokół średniej w zestawie danych. Wyższa wartość odchylenia standardowego oznacza, że wartości są bardziej rozproszone i bardziej różnią się od średniej, podczas gdy niższa wartość odchylenia standardowego wskazuje na mniejsze rozproszenie.

Przykład: wyobraźmy sobie zbiór danych reprezentujący oceny uczniów z egzaminu z matematyki: 80, 85, 90, 75, 95. Wzór, który zostanie przedstawiony poniżej, jest ważny tylko wtedy, gdy pięć wartości, od których zaczęliśmy, tworzy całą populację. Najpierw obliczamy średnią, która wynosi 85. Następnie obliczamy margines zmienności, który wynosi 50.

Najpierw należy obliczyć odchylenia każdego punktu danych od średniej i podnieść wynik każdego z nich do kwadratu:

$(80-85)^2 = (-5)^2 = 25, \quad (85-85)^2 = (0)^2 = 0, \quad (90-85)^2 = (5)^2 = 25, \quad (75-85)^2 = (-10)^2 = 100, (95-85)^2 = (10)^2 = 100$

Wariancja jest średnią tych wartości:

$$\sigma^2 = \frac{25+0+25+100+100}{5} = \frac{250}{5} = 50$$

Na koniec oblicza się odchylenie standardowe, biorąc pierwiastek kwadratowy z marginesu zmienności:

Odchylenie standardowe = $\sqrt{(50)} \approx 7{,}07$

Zatem odchylenie standardowe w tym przypadku wynosi około 7,07. Oznacza to, że średnio wyniki uczniów są oddalone od średniej o około 7,07 jednostki. Odchylenie standardowe jest często wykorzystywane do analizy rozkładu danych i oceny zmienności wartości w zbiorze.

Kwantyle

Kwantyle to wartości, które dzielą uporządkowane rosnąco dane na określone części. Na przykład kwartyle dzielą dane na cztery równe części. Pierwszy kwartyl (Q1) dzieli dolne 25% danych, drugi kwartyl (Q2) jest równy medianie, a trzeci kwartyl (Q3) oddziela górne 25% danych.

Przykład: w zbiorze danych 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, pierwszy kwartyl (Q1) jest równy 6, drugi kwartyl (Q2) jest równy 11, a trzeci kwartyl (Q3) jest równy 16.

1.4 Wyświetlanie statystyk

Prezentacja statystyk obejmuje wykorzystanie różnych metod i narzędzi w celu przedstawienia danych w jasny, przejrzysty i pouczający sposób.

Oto kilka popularnych sposobów wyświetlania statystyk:

Tabele

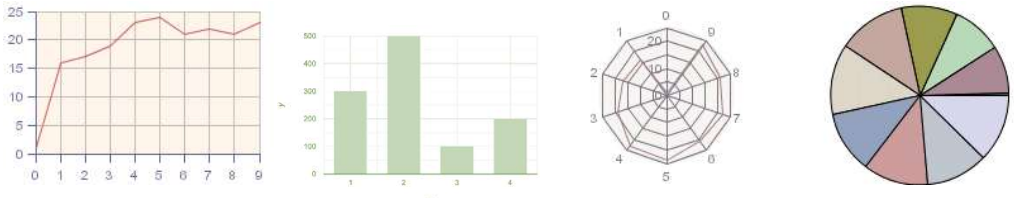
Tabele są podstawową metodą wyświetlania danych. Przykłady obejmują tabele częstotliwości, które pokazują liczbę wystąpień dla różnych wartości, oraz tabele danych, które pokazują więcej informacji o danych.

Oceny uzyskane przez studentów	Liczniki	Częstotliwość
41 - 49	𝍸	3
50 - 58	𝍸 |	6
59 - 67	𝍸	5
68 - 76	𝍸 |	6
77 - 85	||	2
		Razem=22

Rysunek 1.1 Przykładowa tabela częstości

Reprezentacje graficzne

Prezentacje graficzne są skutecznym narzędziem do wizualizacji danych. Obejmują one różne rodzaje wykresów, takie jak wykresy słupkowe, wykresy liniowe, wykresy kołowe, histogramy, wykresy pudełkowe itp.



Rysunek 1.2 Przykłady graficznej reprezentacji danych

Wykresy liniowe służą do wizualizacji trendów i zmian w czasie, dzięki czemu idealnie nadają się do śledzenia danych, które ewoluują w sposób ciągły. Są one szczególnie skuteczne w pokazywaniu relacji między zmiennymi i podkreślaniu wzorców, takich jak wzrosty, spadki lub wahania. Wykresy liniowe są powszechnie stosowane w dziedzinach takich jak finanse, nauka i biznes do analizy danych szeregów czasowych, porównywania trendów w różnych kategoriach lub prognozowania przyszłych zmian na podstawie danych historycznych.

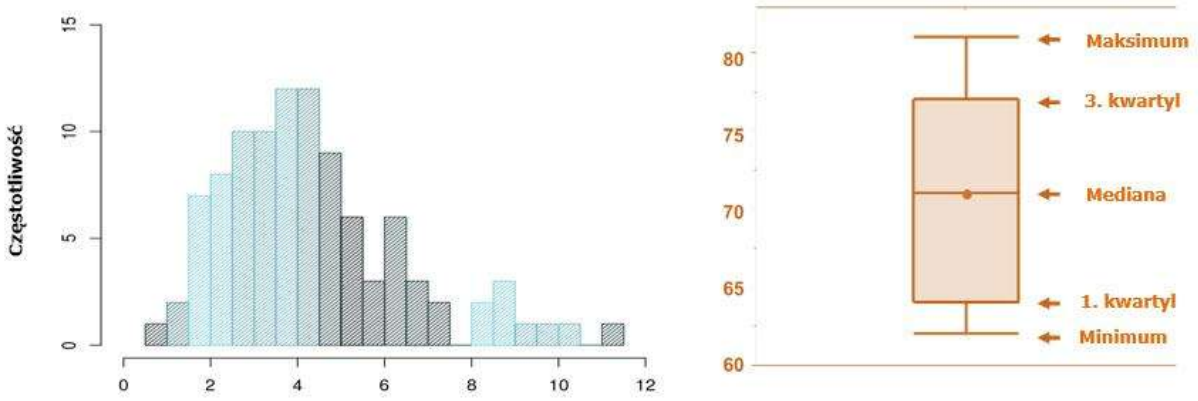
Wykresy słupkowe służą do porównywania ilości w różnych kategoriach, dzięki czemu idealnie nadają się do prezentacji danych dyskretnych. Są one szczególnie skuteczne w podkreślaniu różnic, podobieństw i trendów między grupami. Wykresy słupkowe są powszechnie używane, gdy trzeba pokazać częstotliwości, wartości procentowe lub inne miary liczbowe w jasny i prosty wizualnie sposób. Są szeroko stosowane w biznesie, edukacji i badaniach w celu analizowania i przekazywania kategorialnych.

Wykresy radarowe, znane również jako wykresy pająka, służą do wyświetlania danych wielowymiarowych w wielu wymiarach w formacie kołowym. Są one idealne do porównywania kilku zmiennych lub jednostek pod kątem tych samych kryteriów, podkreślając mocne i słabe strony w jasny, wizualny sposób. Wykresy radarowe są często wykorzystywane w analizie wydajności, podejmowaniu decyzji i porównaniach konkurencji, takich jak ocena cech produktu, umiejętności zespołu lub wyników ankiet w różnych kategoriach.

Wykresy kołowe służą do przedstawiania proporcji lub procentów całości, dzięki czemu idealnie nadają się do wizualizacji względnych rozmiarów różnych kategorii. Są szczególnie skuteczne, gdy chcesz pokazać, w jaki sposób części przyczyniają się do całości lub porównać proporcje na pierwszy rzut oka. Wykresy kołowe są powszechnie używane w raportach, prezentacjach i ankietach do wyświetlania danych, takich jak udział w rynku, alokacja budżetu lub rozkład demograficzny.

Histogramy

Histogramy to graficzne reprezentacje rozkładu danych. Są one używane do pokazania częstotliwości wartości zmiennej w różnych zakresach liczbowych.



Rysunek 1.3 Histogram i wykres kwantylowy (wykres pudełkowy)

Wykres kwantylowy (wykres pudełkowy)

Wykres kwantylowy to rodzaj wykresu stosowanego w statystyce opisowej jako wygodny sposób graficznego przedstawienia grup danych liczbowych poprzez podsumowanie ich pięcioma liczbami: minimum, pierwszy kwartyl, mediana, trzeci kwartyl i maksimum.

Wybór metody wyświetlania statystyk zależy od charakteru danych, celów analizy i grupy docelowej. Ważne jest, aby wybrać metodę, która najlepiej pasuje do przekazu i ułatwia zrozumienie danych.

1.5 Rozkład częstości

Rozkład częstości, znany również jako tabela częstości lub histogram, to sposób na pokazanie liczby wystąpień różnych wartości zmiennej w zestawie danych. Korzystając z rozkładu częstości, można zidentyfikować wzorce, rozkłady i częstotliwości wartości w danych. Jest on powszechnie używany do analizy zmiennych jakościowych (kategorialnych), ale może być również używany do wyświetlania dyskretnych wartości zmiennych ilościowych (numerycznych).

Proces tworzenia rozkładu częstości obejmuje następujące kroki:

- gromadzenie danych: najpierw należy zebrać dane, dla których ma zostać utworzony rozkład częstotliwości,
- zidentyfikowanie różnych wartości: należy zidentyfikować różne wartości, które pojawiają się w danych. Są to kategorie lub wartości dyskretne, które należy przeanalizować,

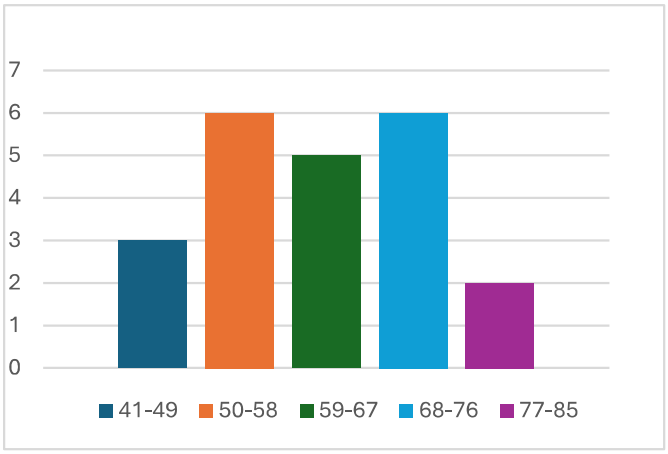
- zliczanie wystąpień: zliczanie, ile razy każda wartość pojawia się w zbiorze danych,
- utworzenie tabeli częstotliwości: należy utworzyć tabelę pokazującą wszystkie różne wartości zmiennej i liczbę wystąpień dla każdej wartości,
- rysowanie histogramu: w przypadku dużej liczby różnych wartości można utworzyć histogram przedstawiający rozkład częstotliwości. Jest to graficzna reprezentacja, która pokazuje liczbę wystąpień dla każdej wartości w postaci słupków.

Przykład rozkładu częstości: Wyobraźmy sobie, że analizowany jest rozkład częstości kolorów samochodów w salonie samochodowym. Zebrane zostały dane dotyczące kolorów 100 samochodów, należy więc sprawdzić, ile samochodów występuje w każdym kolorze?

Tabela 1.1 Nr rozkładu częstości

Punkty uzyskane przez studentów	Liczniki	Częstość
41 - 49	𝍸	3
50 - 58	𝍸 |	6
59 - 67	𝍸	5
68 - 76	𝍸 |	6
77 - 85	||	2
		Razem=22

Wykres rozkładu częstości (histogram) pokazywałby słupki dla każdego koloru z wysokością reprezentującą liczbę samochodów tego koloru. W ten sposób można wyraźnie zobaczyć, który kolor występuje najczęściej i jak rozłożone są inne kolory w zbiorze danych. Rozkłady częstości są przydatnym narzędziem do wizualizacji i analizy danych jakościowych oraz do szybkiego identyfikowania wzorców.



Rysunek 1.4 Wykres rozkładu częstości

1.6 Statystyka opisowa a wnioskowanie

Statystyka opisowa: Statystyka opisowa zajmuje się opisywaniem i podsumowywaniem danych z badanej próby lub populacji. Służy do analizy i zrozumienia danych, ale nie do wnioskowania na temat całej populacji. Głównym celem statystyki opisowej jest opisanie charakterystyki danych, na przykład obliczenie średniej, mediany, zakresu, odchylenia standardowego i stworzenie reprezentacji graficznych, takich jak histogramy lub wykresy. Służy do tworzenia podsumowań i wykresów, które pomagają wizualizować dane.

Statystyka z próby: statystyka inferencyjna zajmuje się wnioskowaniem o populacji na podstawie próby, do której wszystkie jednostki danych miały takie samo prawdopodobieństwo trafienia. Oznacza to, że statystyka wnioskowania pozwala na wyciąganie wniosków na temat całej populacji na podstawie analizy próby reprezentatywnej. Wykorzystuje różne metody statystyczne, takie jak testowanie hipotez, przedziały ufności i analiza regresji, aby zrozumieć, czy zaobserwowane wyniki próby można uogólnić na populację. Na przykład, jeśli chcemy dowiedzieć się, czy średnia wieku w próbie jest reprezentatywna dla całej populacji, użyjemy statystyk inferencyjnych.

Statystyka inferencyjna/matematyczna

Statystyka inferencyjna to gałąź statystyki, która koncentruje się na wnioskach i konkluzjach, jakie możemy wyciągnąć z zebranych danych. Jej głównym zadaniem jest wyciąganie ogólnych wniosków na temat populacji lub próby na podstawie analizy próbki danych.

Główne cele statystyki inferencyjnej to:

Szacowanie parametrów populacji: statystyka inferencyjna pozwala nam oszacować parametry populacji, takie jak średnia, wariancja, proporcje i inne cechy na podstawie próby.

Testowanie hipotez: statystyka inferencyjna może być wykorzystywana do testowania hipotez dotyczących populacji w oparciu o dane z próby. Obejmuje to testy statystyczne, w których porównujemy próbę z założeniami dotyczącymi populacji.

Tworzenie przedziałów ufności: statystyka inferencyjna pozwala nam obliczyć przedziały zawierające szacowane wartości parametrów populacji z określonym poziomem ufności.

Przykład statystyki inferencyjnej: załóżmy, że ma zostać oszacowany średni wzrost wszystkich studentów na uniwersytecie. Ponieważ niemożliwe jest sprawdzenie wszystkich studentów, do rozważań wzięta zostaje próba 100 studentów i to ich wzrost jest mierzony.

Następnie wykorzystujemy statystykę inferencyjną do obliczenia przedziału ufności dla średniego wzrostu wszystkich uczniów. Nasza próba charakteryzuje się średnim wzrostem równym 170 cm i odchyleniem standardowym wynoszącym 5 cm.

Zakładając, że wzrost uczniów w populacji ma w przybliżeniu **rozkład normalny**, można użyć błędu standardowego średniej do obliczenia przedziału ufności. Na przykład, jeśli ma zostać uzyskany 95% przedział ufności, należy użyć błędu standardowego i kwantyli rozkładu normalnego.

Przybliżony 95% przedział ufności dla średniego wzrostu wszystkich studentów na uniwersytecie wynosiłby:

$$170\ cm\ \pm\ 1{,}96\ \times\ (\frac{5\ cm}{\sqrt{100}})\ =\ 170\ cm\ \pm\ 0{,}98\ cm$$

Oznacza to, że można powiedzieć z 95%-ową pewnością, że średni wzrost wszystkich studentów wynosi od około 169,02 cm do 170,98 cm. Ten przedział ufności pozwala nam wywnioskować średni wzrost wszystkich studentów na uniwersytecie z ogólnej próby.

Łącznie opisane powyżej metody statystyczne pozwalają firmom logistycznym lepiej zrozumieć ich procesy, przewidywać przyszłe zdarzenia i podejmować bardziej świadome decyzje w celu poprawy wydajności i konkurencyjności.

1.7 Korelacja i regresja

Są to metody statystyczne wykorzystywane do badania współzależności między wielkościami zmiennymi i przewidywania warunkowej wartości zmiennej zależnej. Obie metody pomagają zrozumieć, na ile jedna zmienna wpływa na drugą i czy pierwsza zmienna może być wykorzystana do przewidywania? Poniżej znajduje się wyjaśnienie każdej z tych dwóch metod:

Korelacja

Korelacja służy do pomiaru stopnia powiązania między dwiema zmiennymi ilościowymi (liczbowymi). Określa, czy istnieje liniowa zależność między dwiema zmiennymi i jak silna jest ta zależność. Korelacja jest mierzona za pomocą współczynnika korelacji, który przyjmuje **wartości od -1** do **1.**

Współczynnik korelacji równy 1 stanowi doskonałą korelację dodatnią, co oznacza, że zmienne są doskonale skorelowane i poruszają się w tym samym kierunku.

Współczynnik korelacji równy -1 oznacza idealną korelację ujemną, co oznacza, że dwie zmienne są całkowicie odwrotnie skorelowane i poruszają się w przeciwnych kierunkach.

Współczynnik korelacji równy 0 oznacza, że nie ma liniowej zależności między zmiennymi.

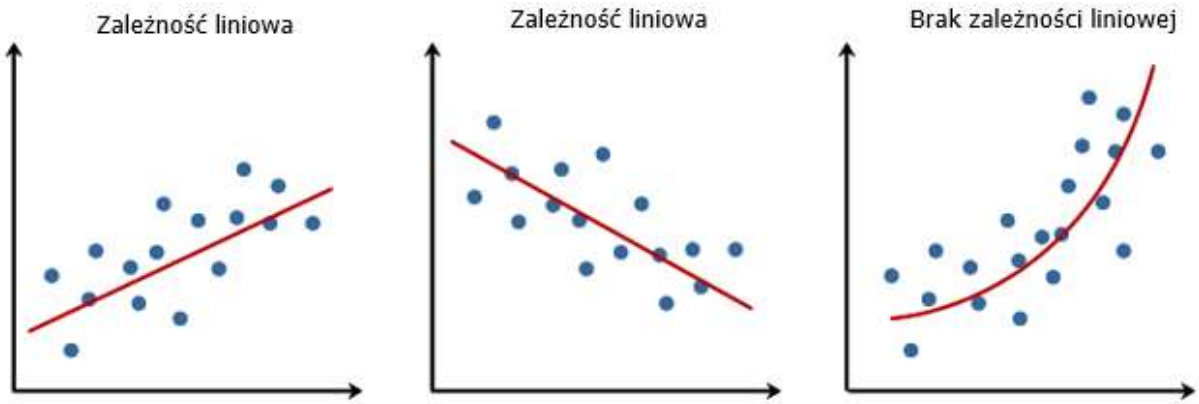
Przykład: korelacja między liczbą godzin nauki a ocenami uzyskiwanymi przez studentów będzie dodatnia, jeśli wzrost liczby godzin nauki zwykle odpowiada wyższym ocenom.

Regresja

Regresja służy do modelowania i przewidywania wartości jednej zmiennej ilościowej (zmiennej zależnej) na podstawie wartości innej zmiennej ilościowej (zmiennej niezależnej). Istnieją różne rodzaje regresji, w tym **prosta regresja liniowa, wieloraka regresja liniowa**, regresja logistyczna itp.

Regresja liniowa prosta: służy do modelowania związku między jedną zmienną niezależną a jedną zmienną zależną. Model jest liniowy i jest zwykle reprezentowany przez równanie linii prostej ($y = a + bx$), gdzie a jest punktem przecięcia z osią y a b jest nachyleniem linii.

Wieloraka regresja liniowa: używana, gdy należy modelować związek między kilkoma zmiennymi niezależnymi a jedną zmienną zależną.



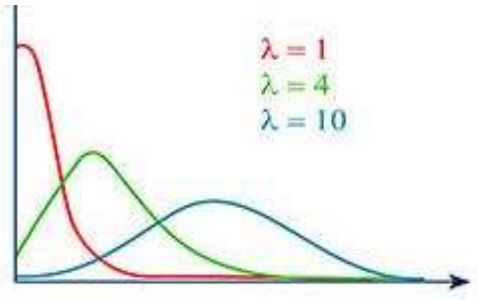
Rysunek 1.5 Wykresy prostej regresji liniowej

Przykład: regresja liniowa prosta może być wykorzystana do modelowania związku między liczbą ukończonych zadań edukacyjnych (zmienna niezależna) a oceną z egzaminu końcowego (zmienna zależna).

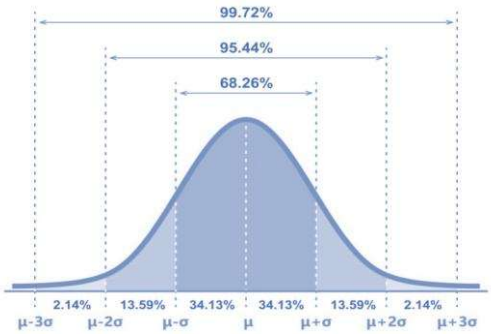
1.8 Rozkłady prawdopodobieństwa

W statystyce rozkład prawdopodobieństwa opisuje prawdopodobieństwa różnych wartości, jakie może przyjąć zmienna. Jest to model matematyczny, który pomaga nam zrozumieć i analizować zjawiska losowe oraz przewidywać rozkład wartości w określonych okolicznościach. Istnieje kilka różnych rozkładów prawdopodobieństwa, z których każdy ma swoją własną charakterystykę i zastosowanie w różnych sytuacjach. Oto niektóre z najbardziej znanych rozkładów prawdopodobieństwa w statystyce:

Rozkład normalny (Gaussa): rozkład normalny jest jednym z najważniejszych i najszerzej stosowanych rozkładów. Opisuje on symetryczny rozkład w kształcie dzwonu ze znanymi parametrami: średnią (μ) i odchyleniem standardowym (σ). Wiele naturalnych zjawisk jest zbliżonych do rozkładu normalnego.



Rysunek 1.6 Wykres rozkładu Poissona



Rysunek 1.7 Wykres rozkładu normalnego

Rozkład dwumianowy: rozkład dwumianowy jest używany do modelowania liczby sukcesów (np. liczby wyrzuconych „orzełków") w danej liczbie niezależnych prób Bernoulliego. Ma on dwa parametry: liczbę prób (n) i prawdopodobieństwo sukcesu(p).

Rozkład Poissona: rozkład Poissona służy do modelowania liczby zdarzeń występujących w danym okresie czasu lub przestrzeni. Jest on zwykle używany do modelowania rzadkich zdarzeń, takich jak wypadki, wezwania służb ratunkowych itp. Parametrem rozkładu jest średnia liczba wystąpień zdarzenia(λ).

Rozkład wykładniczy: rozkład wykładniczy jest szczególnym przypadkiem rozkładu gamma i jest używany do modelowania czasów do pierwszego zdarzenia w procesie Poissona. Parametrem rozkładu jest średnia (λ).

Rozkład t-Studenta: rozkład t-Studenta jest używany do szacowania przedziałów ufności i testowania hipotez, gdy masz małą próbkę i nie znasz odchylenia standardowego populacji. Jest to ważne podczas analizy próbek, w przypadku których założenie o rozkładzie normalnym nie ma uzasadnienia.

Rozkład chi-kwadrat: rozkład chi-kwadrat jest używany do analizy rozkładu częstotliwości w tabelach, testowania niezależności i testowania hipotez. Jest on często wykorzystywany w testach statystycznych, takich jak test chi-kwadrat.

Rozkład F: rozkład F jest używany do porównywania zmienności między dwiema próbkami. Jest używany w analizie wariancji (ANOVA) i innych testach statystycznych.

Te rozkłady prawdopodobieństwa są podstawowymi elementami składowymi w statystyce i są wykorzystywane do modelowania i analizowania różnych typów danych w różnych

kontekstach. Wybór właściwego rozkładu prawdopodobieństwa ma kluczowe znaczenie podczas przeprowadzania analiz statystycznych i przewidywania wyników.

BIBLIOGRAFIA DO ROZDZIAŁU 1.

- Introductory Statistics. Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/9789815123135123010l
- Introductory Statistics 2e, Openstax, Rice University, Houston, Texas 77005, Jun 23, senior contributing authors: Barbara Illowsky and Susan dean, De anza college, Publish Date: Dec 13, 2023, (https://openstax.org/details/books/introductory-statistics-2e);
- Introductory Statistics 4th Edition, Susan Dean and Barbara Illowsky, Adapted by Riyanti Boyd & Natalia Casper (Published 2013 by OpenStax College) July 2021, (http://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf)
- Introductory Statistics 7th Edition, Prem S. Mann, eastern Connecticut state university with the help of Christopher Jay Lacke, Rowan university, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, 2011
- Journal of the Royal Statistical Society 2024, A reputable journal publishing cutting-edge research and articles on various aspects of statistics, including theoretical advancements and practical applications. Recent issues have featured studies on sampling and hypothesis testing
- Introduction to statistics, made easy second edition, Prof. Dr. Hamid Al-Oklah Dr. Said Titi Mr. Tareq Alodat, March 2014
- Statistics for Business and Economics, Thirteenth Edition, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, 2017, 2015 Cengage Learning®
- Statistics for Business, First edition, Derek L Waller, 2008 Copyright © 2008, Derek L Waller, Published by Elsevier Inc. All rights reserved

Dodatkowe linki do literatury i filmów na Youtube do Rozdziału 1

https://open.umn.edu/opentextbooks/textbooks/196

https://www.scribbr.com/category/statistics/

https://stats.libretexts.org/Bookshelves/Introductory_Statistics

https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf

https://saylordotorg.github.io/text_introductory-statistics/

https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20(7th%20Ed).pdf

https://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf

https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/

https://onlinestatbook.com/Online_Statistics_Education.pdf

https://www.researchgate.net/profile/Tareq-Alodat-2/publication/340511098_INTRODUCTION_TO_STATISTICS_MADE_EASY/links/5e8de3dc4585150839c7b58a/INTRODUCTION-TO-STATISTICS-MADE-EASY.pdf

https://byjus.com/maths/statistics/

https://www.khanacademy.org/math/statistics-probability

https://www.youtube.com/watch?v=XZo4xyJXCak

https://www.youtube.com/watch?v=LMSyiAJm99g

https://www.youtube.com/watch?v=VPZD_aij8H0

https://www.youtube.com/watch?v=TLwp5DwcqD4

https://www.youtube.com/watch?v=fpFj1Re1l84

https://youtube.com/playlist?list=PLqzoL9-eJTNAB5st3mtP_bmXafGSH1Dtz&si=z-IXQ1iKbw2-ieJW

https://www.youtube.com/watch?v=44MJyNTxaP8