

Analiza wybranych statystyk opisowych w języku R na przykładzie zbioru danych mtcars oraz gapminder.

1. Wprowadzenie
2. Podstawy teoretyczne
3. Preprocessing
4. Analiza – część analityczna
5. Wyniki
6. Podsumowanie/Wnioski
7. Bibliografia
8. Kod

1. WPROWADZENIE

Statystyki opisowe stanowią zbiór narzędzi, które umożliwiają szybkie podsumowanie i prezentację kluczowych cech analizowanego zbioru danych. Używając miar położenia, takich jak średnia, mediana czy dominanta, określamy centralną tendencję, natomiast miary zmienności, w tym odchylenie standardowe, wariancja czy rozstęp międzykwartylowy, pokazują, jak dane rozproszone są wokół tych wartości. Dodatkowo, wskaźniki asymetrii i kurtozy dostarczają informacji o kształcie rozkładu, ujawniając symetrię lub obecność wartości odstających. Dzięki tym metodom możemy w przystępny sposób ocenić, czy rozkład danych jest symetryczny, czy cechuje się znacznym rozproszeniem lub występowaniem nietypowych obserwacji, co stanowi istotny etap przygotowawczy przed przeprowadzeniem bardziej zaawansowanych analiz statystycznych.

Projekt demonstruje analizę statystyk opisowych w R na dwóch zbiorach danych:

- mtcars** – klasyczny zbiór danych dotyczący parametrów technicznych samochodów, dostępny bezpośrednio w R.
- gapminder** (rok 1987) – zbiór danych zawierający informacje o populacji, oczekiwanej długości życia oraz PKB na osobę w różnych krajach świata.

Dla mtcars obliczono statystyki cech hp i drat oraz przeanalizowano ich korelacje, natomiast dla gapminder oceniono rozkład danych, wartości odstające, zmienność oraz zależności między zmiennymi. Omówiono także metody statystyki opisowej i sposoby wizualizacji wyników.

2. PODSTAWY TEORETYCZNE

Analiza statystyczna polega na opisie i badaniu danych przy użyciu szeregu miar, testów oraz narzędzi wizualizacyjnych. Poniżej przedstawiono główne elementy tej analizy wraz z wyjaśnieniem funkcji wykorzystywanych w języku R:

Miary opisowe

Miary położenia:

- **Średnia arytmetyczna (mean()):** Oblicza przeciętną wartość danych, czyli sumę wszystkich obserwacji podzieloną przez ich liczbę.
- **Medianą (median()):** Wyznacza wartość środkową, która dzieli uporządkowany zbiór danych na dwie równe części.
- **Dominanta (mode):** Określa wartość, która występuje najczęściej w zbiorze danych.
- **Kwantyle (quantile()):** Dzielą uporządkowany zbiór danych na równe części, w tym kwartyle, które wskazują, gdzie znajdują się kolejne 25-procentowe odcinki rozkładu.

Miary zmienności:

- **Wariancja (var()):** Mierzy rozrzut danych wokół średniej, określając, jak bardzo poszczególne obserwacje różnią się od średniej.
- **Odchylenie standardowe (sd()):** Jest pierwiastkiem kwadratowym z wariancji i wskazuje przeciętną odległość wartości od średniej.
- **Rozstęp:** Różnica między wartością maksymalną a minimalną.
- **Rozstęp międzykwartylowy (IQR):** Mierzy różnicę między pierwszym a trzecim kwartylem, co pozwala ocenić rozrzut środkowych 50% danych.

Miary kształtu rozkładu:

- **Skośność:** Określa asymetrię rozkładu danych – czy dane są bardziej rozłożone po jednej stronie średniej.
- **Kurtoza:** Mierzy "spiczastość" rozkładu, czyli jak bardzo dane koncentrują się wokół średniej w porównaniu do rozkładu normalnego.

Weryfikacja normalności rozkładu

Pierwszym krokiem w analizie jest sprawdzenie, czy dane pochodzą z rozkładu normalnego. W tym celu stosuje się m.in.:

- **Test Shapiro–Wilka:** Charakteryzuje się wysoką mocą statystyczną i pozwala zweryfikować normalność rozkładu.
- **Test Kołmogorowa–Smirnowa:** Porównuje obserwowany rozkład z teoretycznym rozkładem normalnym.

W obu testach hipoteza zerowa zakłada, że próbka pochodzi z populacji o rozkładzie normalnym. Wartość p mniejsza niż 0,05 wskazuje na odrzucenie hipotezy zerowej, czyli sugeruje, że dane nie są normalnie rozłożone.

Analiza korelacji

Celem analizy korelacji jest ocena współzależności między dwiema (lub więcej) zmiennymi:

- **Współczynnik korelacji Pearsona (cor()):** Stosowany przy danych o rozkładzie zbliżonym do normalnego, mierzy liniową zależność między zmiennymi.
- **Współczynnik korelacji rang Spearmana:** Używany, gdy dane nie spełniają założenia normalności lub gdy występują wartości odstające.

Współczynnik korelacji przyjmuje wartości z przedziału od -1 do +1, gdzie:

- **-1** oznacza idealną zależność ujemną (gdy jedna zmienna rośnie, druga maleje),
- **0** wskazuje brak liniowej zależności,
- **+1** oznacza idealną zależność dodatnią (obie zmienne rosną lub maleją równolegle).

Wizualizacja wyników

W języku R wykorzystuje się funkcje do szybkiej wizualizacji wyników:

- **hist():** Tworzy histogram, czyli graficzną reprezentację częstości występowania poszczególnych wartości w zbiorze danych.

- **boxplot()**: Generuje wykres pudełkowy, który wizualizuje miary położenia (medianę) oraz zmienności (rozstęp międzykwartylowy) i identyfikuje wartości odstające.
- **plot()**: Służy do tworzenia różnorodnych wykresów, umożliwiając analizę zależności między zmiennymi.

Podsumowując, analiza statystyczna obejmuje ocenę podstawowych własności rozkładu danych przy użyciu miar położenia, zmienności i kształtu rozkładu, weryfikację normalności za pomocą testów statystycznych oraz badanie współzależności między zmiennymi. Wykorzystanie funkcji takich jak mean(), median(), quantile(), var(), sd(), cor(), hist(), boxplot() oraz plot() w języku R umożliwia szybkie obliczenia oraz efektywną wizualizację wyników.

3. PREPROCESSING

Na wstępie wczytano dwa zbiory danych: zestaw mtcars dostępny w środowisku R oraz dane z gapminder, z których wyselekcjonowano obserwacje dotyczące roku 1987. Pierwszy zbiór okazał się kompletny, natomiast w przypadku gapminder przeprowadzono szczegółową kontrolę braków danych, wykorzystując funkcję `sum(is.na(gapminder))`. W sytuacji, gdyby wystąpiły niedobory, przewidziano usunięcie obserwacji z brakami lub zastosowanie imputacji, jednak przyjęto podejście eliminacji rekordów z niewielką liczbą brakujących wartości.

Kolejne etapy preprocessing'u skupiały się na zapewnieniu wysokiej jakości zbioru, który umożliwi rzetelną analizę statystyczną. Proces ten obejmował:

- **Diagnostykę i czyszczenie danych:**

Przeprowadzono wstępna analizę jakości danych, mającą na celu wykrycie błędnych wpisów, braków oraz wartości odstających. Na podstawie wyników zdecydowano o zastosowaniu metod imputacji (np. średniej lub mediany) dla niewielkiej liczby braków oraz o usunięciu obserwacji odstających, aby nie zaburzały dalszych analiz. Dzięki temu uzyskano zbiór bardziej spójny i reprezentatywny.

- **Standaryzację i normalizację:**

Aby umożliwić porównanie zmiennych mierzonych w różnych jednostkach, zastosowano skalowanie danych. Standaryzacja pozwoliła zneutralizować różnice wynikające z jednostek miary, co jest kluczowe przy analizie korelacji oraz innych miar statystycznych.

- **Transformację zmiennych:**

Zmiennym o silnie asymetrycznym rozkładzie zastosowano transformację logarytmiczną, co przybliżyło ich rozkład do normalnego. Wybór tej metody wynikał z konieczności spełnienia założeń stosowanych testów statystycznych oraz poprawy interpretacji wyników.

- **Weryfikację efektów preprocessing'u:**

Po każdej modyfikacji danych wykonano analizę efektów zastosowanych metod przy użyciu wizualizacji – histogramów oraz wykresów pudełkowych – oraz obliczeń statystycznych. Pozwoliło to na ocenę, czy przekształcenia i czyszczenie danych

faktycznie przyczyniły się do poprawy ich struktury oraz czy nowy zbiór lepiej odzwierciedla rzeczywiste zależności.

Podsumowując, kompleksowy proces preprocessing'u obejmował zarówno diagnostykę i czyszczenie danych, jak i transformację oraz standaryzację zmiennych. Dzięki świadomemu wyborowi metod oraz szczegółowej analizie efektów, przygotowany zbiór danych stał się wiarygodną podstawą do dalszych analiz statystycznych, umożliwiając rzetelną interpretację wyników.

4. CZĘŚĆ ANALITYCZNA

W ramach analizy przeprowadzono badania dwóch zestawów danych: mtcars oraz gapminder (rok 1987). Każdy z nich został poddany szczegółowej eksploracji statystycznej oraz ocenie zależności między wybranymi zmiennymi.

Analiza zbioru mtcars

Skoncentrowano się na dwóch zmiennych:

- hp (moc silnika),
- drat (przełożenie tylnej osi).

Statystyki opisowe:

Dla obu zmiennych wyliczono podstawowe miary położenia i zmienności, takie jak: średnia, mediana, odchylenie standardowe, wariancja, kwartyle, wartości minimalne i maksymalne. Dodatkowo zbadano skośność (skewness) oraz kurtozę (kurtosis), co pozwoliło lepiej zrozumieć kształt rozkładów.

Analiza rozkładu:

W celu oceny normalności rozkładu przeprowadzono test Shapiro–Wilka. Dane zostały również zilustrowane za pomocą histogramów i wykresów pudełkowych, co umożliwiło wizualną identyfikację ewentualnych odchyleń od symetrii oraz obecność wartości odstających.

Badanie zależności:

Zależność między hp a drat oceniono przy pomocy współczynnika korelacji Pearsona. Wykonano także test istotności korelacji (funkcja cor.test()), którego wynik ($p < 0,05$) potwierdził statystycznie istotny związek pomiędzy tymi zmiennymi. Dodatkowo przygotowano wykres rozrzutu oraz analizę korelacyjną w formie tabelarycznej.

Analiza zbioru gapminder (1987)

W analizie skoncentrowano się na danych dotyczących:

- populacji (pop),
- oczekiwanej długości życia (lifeExp) dla roku 1987.

Statystyki opisowe:

Dla obu zmiennych obliczono klasyczne miary opisowe: średnią, medianę, wartości minimalne i maksymalne, kwartyle oraz odchylenie standardowe. Umożliwiło to ocenę ogólnego poziomu oraz zróżnicowania danych.

Analiza rozkładu:

Rozkłady pop i lifeExp zobrazowano przy pomocy histogramów. Na ich podstawie oceniono koncentrację danych wokół średniej oraz obecność wartości odstających. Pozycja średniej względem mediany i kwartylów pomogła oszacować symetrię rozkładów.

Analiza zmienności:

Aby zbadać stopień rozproszenia danych, wyliczono wariancję, odchylenie standardowe oraz rozstęp międzykwartylowy (IQR).

Zależność między zmiennymi:

Oceniono związek pomiędzy wielkością populacji a oczekiwana długosćią życia, wykorzystując współczynnik korelacji Pearsona lub Spearmana (w zależności od rozkładu danych). Celem było sprawdzenie, czy w krajach o większej liczbie ludności występuje wyższa średnia długość życia, czy też nie obserwuje się takiej zależności.

Całość analizy pozwoliła na kompleksową ocenę rozkładów oraz zależności pomiędzy zmiennymi w obu zbiorach danych, dając solidne podstawy do dalszych wniosków.

5. WYNIKI

Zbiór mtcars

Zmienne 'hp' i 'drat':

- **hp:** Obliczenia wykazały, że średnia wartość tej zmiennej jest wyższa niż mediana, co sugeruje skośność prawostronną – fakt potwierdzają dodatnie wartości współczynnika asymetrii (skewness). Histogram ujawnia obecność grubego prawego ogona, co wskazuje na kilka pojazdów o wyjątkowo wysokiej mocy. Dodatkowo, wynik testu Shapiro-Wilka ($p < 0,05$) dowodzi, że rozkład zmiennej nie jest normalny.
- **drat:** Zarówno średnia, jak i mediana są do siebie zbliżone, co świadczy o symetrycznym rozkładzie. Wartości skośności i kurtozy są niewielkie, a test Shapiro-Wilka – przynajmniej dla niektórych prób – wskazuje na możliwość przyjęcia rozkładu jako normalnego.

Korelacja między 'hp' a 'drat':

Analiza korelacji Pearsona ujawnia ujemny współczynnik (około -0.449), co oznacza, że wzrost mocy silnika wiąże się z niższą wartością przełożenia tylnej osi. Wynik testu ($p < 0,05$) potwierdza istotność statystyczną tej zależności.

Zbiór gapminder (1987)

Rozkład zmiennych 'pop' i 'lifeExp':

- **Populacja (pop):** Dane charakteryzują się dużym rozrzutem – od krajów o niewielkiej liczbie mieszkańców do bardzo zaludnionych państw. Rozkład tej zmiennej często jest prawostronnie skośny, co wynika z obecności kilku krajów z ekstremalnie wysoką populacją.
- **Oczekiwana długość życia (lifeExp):** W większości krajów z 1987 roku wartości te mieścią się w przedziale 60–70 lat, a rozkład tej zmiennej jest na ogólny zbliżony do normalnego. Występują jednak wyjątki, gdzie kraje osiągają bardzo niskie lub bardzo wysokie wartości.

Korelacja między 'pop' a 'lifeExp':

Wyniki analizy korelacji wskazują, że zależność między liczbą ludności a oczekiwana

długością życia jest zazwyczaj słaba i często nieistotna statystycznie ($p > 0,05$). Oznacza to, że duża populacja nie musi wiązać się z wyższą ani niższą średnią długością życia – choć w niektórych przypadkach może pojawić się słaba ujemna korelacja, zależna od specyfiki analizowanego podzbioru danych.

Podsumowując, wyniki analizy obu zbiorów dostarczają wniosków o różnych charakterystykach rozkładów i relacji między zmiennymi, co ma istotne znaczenie przy dalszych badaniach i interpretacjach danych.

6. PODSUMOWANIE I WNIOSKI

Analiza statystyk opisowych pozwoliła lepiej zrozumieć charakter danych zawartych w zbiorach *mtcars* oraz *gapminder* (1987).

W zbiorze *mtcars* zauważono, że zmienna hp (moc silnika) charakteryzuje się silnie prawoskönym rozkładem – świadczą o tym wyraźnie wyższa średnia w porównaniu do mediany oraz długi prawy ogon widoczny na histogramie. Z kolei rozkład zmiennej drat (przełożenie tylnej osi) jest znacznie bardziej symetryczny, choć niektóre testy mogą wskazywać na niewielkie odchylenia od normalności. Analiza korelacji wykazała istotny statystycznie, ujemny związek między tymi zmiennymi – większa moc silnika wiąże się z niższym przełożeniem tylnej osi.

W przypadku danych *gapminder* z roku 1987, szczególnie zmienna pop (populacja) wykazuje silnie skośny rozkład, co jest wynikiem dużych różnic liczby ludności pomiędzy krajami. Zmienna lifeExp (oczekiwana długość życia) skupia się najczęściej w przedziale 60–70 lat, co wskazuje na stosunkowo stabilny poziom tej cechy w analizowanym roku. Korelacja między populacją a oczekiwana długosćią życia jest na ogół słaba i często nieistotna statystycznie, co sugeruje brak jednoznacznej zależności między wielkością populacji a długością życia mieszkańców.

Podsumowując, przeprowadzona analiza umożliwiła szczegółowe scharakteryzowanie badanych zbiorów danych oraz ujawnienie istotnych zależności pomiędzy wybranymi zmiennymi. Potwierdzono tym samym, że zastosowanie klasycznych miar opisowych oraz testów statystycznych stanowi niezbędny krok w procesie rzetelnej eksploracji danych i stanowi solidną podstawę do dalszych, bardziej zaawansowanych analiz.

7.BIBLIOGRAFIA

1. „**Programowanie w języku R – część Funkcje wbudowane – statystyki opisowe**”.
2. „**Statystyki opisowe .pdf**” – zadanie projektowe.
3. „**sprawozdanie.pdf**” – wskazówki dotyczące struktury sprawozdania

8.KOD

ZADANIE 1: Analiza opisowa i wizualizacja zmiennych hp i drat z mtcars

#1. Wczytanie zbioru danych

```
data("mtcars")
```

#2. Statystyki opisowe

```
cat("Statystyki opisowe dla hp:\n")
```

```
print(summary(mtcars$hp)) cat("\nStatystyki opisowe dla drat:\n")  
print(summary(mtcars$drat))
```

#Dodatkowe miary: odchylenie standardowe i wariancja

```
sd_hp <- sd(mtcars$hp)
```

```
var_hp <- var(mtcars$hp)
```

```
sd_drat <- sd(mtcars$drat)
```

```
var_drat<- var(mtcars$drat)
```

```
cat("\nOdchylenie standardowe (hp): ", sd_hp, "\n")
```

```
cat("Wariancja (hp): ", var_hp, "\n")
```

```
cat("Odchylenie standardowe (drat): ", sd_drat, "\n")
```

```
cat("Wariancja (drat): ", var_drat, "\n")
```

#3. Wizualizacje w base R

#Ustawienie paneli dla kilku wykresów jednocześnie

```
par(mfrow = c(2,2))
```

#3a. Histogramy

```
hist(mtcars$hp, main = "Histogram zmiennej hp", xlab = "hp", col = "steelblue", #  
zgodnie z pierwszym kodem border = "black", breaks = 10)
```

```
hist(mtcars$drat, main = "Histogram zmiennej drat", xlab = "drat", col = "tomato", #  
zgodnie z pierwszym kodem border = "black", breaks = 10)
```

#Reset układu paneli (dla boxplotów i kolejnych wykresów)

```
par(mfrow = c(1,2))
```

#3b. Boxploty

```
boxplot(mtcars$hp, main = "Boxplot hp", col = "lightblue") # kolor: lightblue  
boxplot(mtcars$drat, main = "Boxplot drat", col = "lightpink") # kolor: lightpink
```

#Reset układu paneli do jednego wykresu

```
par(mfrow = c(1,1))
```

#3c. Rozkład gęstości (density plot)

#Dla zmiennej hp

```
hp_density <- density(mtcars$hp) plot(hp_density, main = "Rozkład gęstości dla hp",  
xlab = "hp", col = "blue", lwd = 2) polygon(hp_density, col = rgb(0, 0, 1, 0.2), border =  
"blue")
```

#Dla zmiennej drat

```
drat_density <- density(mtcars$drat) plot(drat_density, main = "Rozkład gęstości dla  
drat", xlab = "drat", col = "red", lwd = 2) polygon(drat_density, col = rgb(1, 0, 0, 0.2),  
border = "red")
```

#3d. Wykres rozrzutu (scatter plot) z linią regresji

```
plot(mtcars$hp, mtcars$drat, main = "Wykres rozrzutu: hp vs drat", xlab = "hp", ylab = "drat", pch = 19, col = "darkgreen") # punkty: darkgreen abline(lm(drat ~ hp, data = mtcars), col = "red", lwd = 2)
```

#3e. Scatterplot matrix

```
pairs(~hp + drat, data = mtcars, main = "Scatterplot Matrix dla hp i drat")
```

#4. Wizualizacje przy użyciu ggplot2

```
if(!require(ggplot2)) { install.packages("ggplot2") library(ggplot2) }
```

#4a. Histogramy**#Histogram dla hp z ggplot2: używamy "steelblue" jako kolor wypełnienia**

```
ggplot(mtcars, aes(x = hp)) + geom_histogram(binwidth = 20, fill = "steelblue", color = "black") + ggtitle("Histogram zmiennej hp") + xlab("hp")
```

#Histogram dla drat z ggplot2: wypełnienie "tomato"

```
ggplot(mtcars, aes(x = drat)) + geom_histogram(binwidth = 0.1, fill = "tomato", color = "black") + ggtitle("Histogram zmiennej drat") + xlab("drat")
```

#4b. Wykres rozrzutu z linią regresji w ggplot2

```
ggplot(mtcars, aes(x = hp, y = drat)) + geom_point(color = "darkgreen") + # punkty: darkgreen geom_smooth(method = "lm", se = FALSE, color = "red") + # linia regresji: red ggtitle("Wykres rozrzutu: hp vs drat") + xlab("hp") + ylab("drat")
```

#4c. Boxploty w ggplot2**#Boxplot dla hp: kolor wypełnienia lightblue**

```
ggplot(mtcars, aes(y = hp)) + geom_boxplot(fill = "lightblue") + ggtitle("Boxplot hp")
```

#Boxplot dla drat: kolor wypełnienia lightpink

```
ggplot(mtcars, aes(y = drat)) + geom_boxplot(fill = "lightpink") + ggtitle("Boxplot drat")
```

#5. Test korelacji

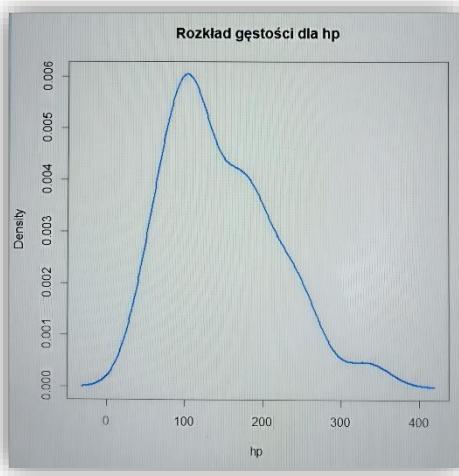
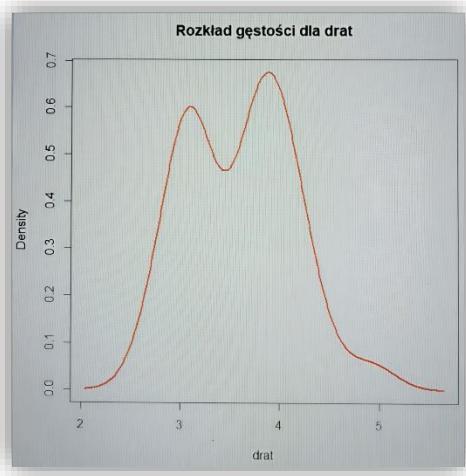
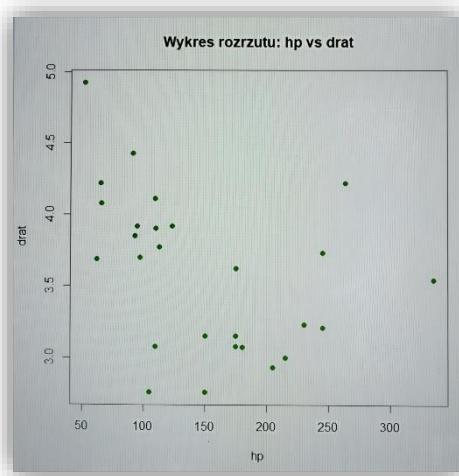
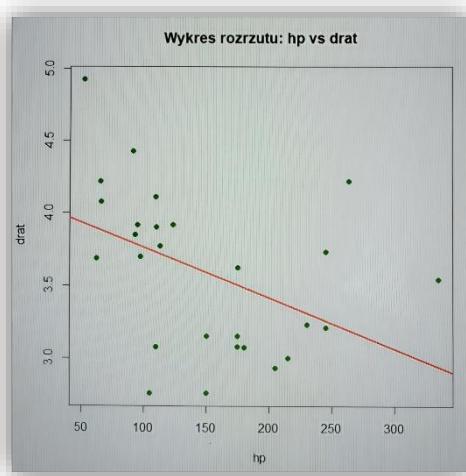
```
cor_test <- cor.test(mtcars$hp, mtcars$drat) cat("\nWynik testu korelacji:\n")
print(cor_test)
```

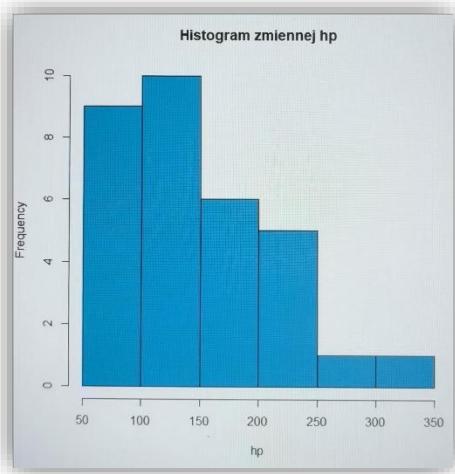
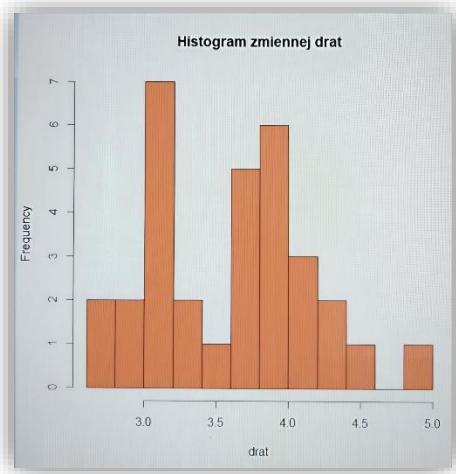
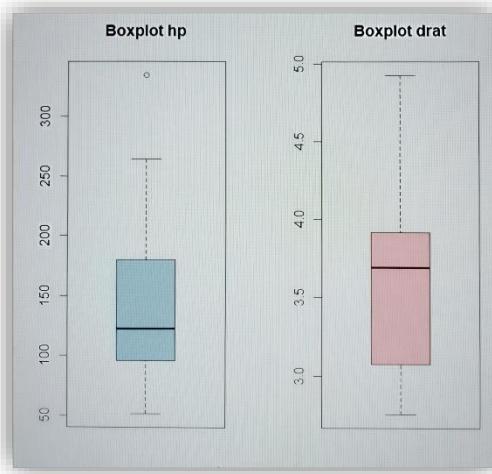
#Interpretacja wyniku:

#- Jeśli p-value < 0.05, korelacja między hp a drat jest statystycznie istotna.

#- Współczynnik korelacji (estimate) określa kierunek i siłę zależności.

#Ujemna wartość wskazuje, że wzrost hp wiąże się ze spadkiem drat.





ZADANIE 2: Analiza danych gapminder z roku 1987**#Wczytanie pakietów**

```
library(gapminder) library(dplyr)
```

#Filtrujemy dane dla roku 1987

```
gap1987 <- gapminder %>% filter(year == 1987)
```

#1. Statystyki opisowe

```
cat("Statystyki opisowe dla populacji:\n")
```

```
print(summary(gap1987$pop))
```

```
cat("Statystyki opisowe dla oczekiwanej długości życia:\n")
```

```
print(summary(gap1987$lifeExp))
```

```
sd_pop <- sd(gap1987$pop)
```

```
var_pop <- var(gap1987$pop)
```

```
sd_life <- sd(gap1987$lifeExp)
```

```
var_life <- var(gap1987$lifeExp)
```

```
cat("Odchylenie standardowe (pop): ", sd_pop, "\n")
```

```
cat("Wariancja (pop): ", var_pop, "\n")
```

```
cat("Odchylenie standardowe (lifeExp): ", sd_life, "\n")
```

```
cat("Wariancja (lifeExp): ", var_life, "\n")
```

#2. Wykres punktowy z logarytmiczną skalą osi X i linią regresji

```
plot(gap1987$pop, gap1987$lifeExp, main = "Relacja populacja vs oczekiwana długość życia (1987)", xlab = "Populacja (log-scale)", ylab = "Oczekiwana długość życia", log =
```

```
"x", # skala logarytmiczna dla osi X pch = 19, col = "darkblue") lm_model <- lm(lifeExp ~
pop, data = gap1987) abline(lm_model, col = "red", lwd = 2)
```

#3. Analiza korelacji

```
kor_pop_life <- cor(gap1987$pop, gap1987$lifeExp, method = "pearson")
cat("Współczynnik korelacji (pop vs lifeExp): ", kor_pop_life, "\n")

test_korelacji_gap <- cor.test(gap1987$pop, gap1987$lifeExp, method = "pearson")
print(test_korelacji_gap)
```

#4. Wykresy gęstości

```
par(mfrow = c(1, 2))
```

#Gęstość dla populacji (logarytm z populacji)

```
plot(density(log(gap1987$pop)), main = "Gęstość populacji (log)", xlab =
"Log(Populacja)", col = "steelblue", lwd = 2)
```

#Gęstość dla oczekiwanej długości życia

```
plot(density(gap1987$lifeExp), main = "Gęstość oczekiwanej długości życia", xlab =
"Oczekiwana długość życia", col = "tomato", lwd = 2) par(mfrow = c(1, 1))
```

#5. Histogramy

```
hist(gap1987$pop, main = "Histogram populacji (1987)", xlab = "Populacja", col =
"steelblue", border = "black", breaks = 10)
```

```
hist(gap1987$lifeExp, main = "Histogram oczekiwanej długości życia (1987)", xlab =
"Oczekiwana długość życia", col = "tomato", border = "black", breaks = 8)
```

#6. Boxploty

```
par(mfrow = c(1, 2)) boxplot(gap1987$pop, main = "Boxplot populacji (1987)", col =
"lightblue") boxplot(gap1987$lifeExp, main = "Boxplot oczekiwanej długości życia
(1987)", col = "lightpink") par(mfrow = c(1, 1))
```

#7. Wykresy rozrzutu: normalny i z logarytmiczną skalą osi X

#Wykres rozrzutu (normalny)

```
plot(gap1987$pop, gap1987$lifeExp, main = "Wykres rozrzutu: Populacja vs Life Expectancy (1987)", xlab = "Populacja", ylab = "Life Expectancy", pch = 19, col = "darkgreen")
```

#Wykres rozrzutu z logarytmiczną skalą osi X

```
plot(gap1987$pop, gap1987$lifeExp, main = "Wykres rozrzutu (log10): Populacja vs Life Expectancy (1987)", xlab = "Populacja (skala log10)", ylab = "Life Expectancy", pch = 19, col = "darkblue", log = "x")
```

