

Projekt 2

Wstęp do Analizy Danych | Politechnika Krakowska

Jakub Kapała

Numer albumu: 151885

Data: 15.05.2025

Treść projektu

1. Regresja liniowa w R wykorzystuje funkcję `lm()` do stworzenia modelu regresji. Aby przyjrzeć się modelowi, użyj funkcji `summary()`. Wybierz trzy dowolne próbki z biblioteki `MASS` i dokonaj analizy regresji przeprowadzając diagnostykę modelu regresji na podstawie wykresów diagnostycznych w R. Proszę opisać co oznaczają te wykresy i w jaki sposób można je wykorzystać do sprawdzenia czy założenia modelu są spełnione. Funkcja: `plot(model, which = 1:4)`.
2. Znajdź współczynniki regresji (i) minimalizując sumę kwadratów reszt - metoda najmniejszych kwadratów.

$$\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - (b_0 + b_1 X_i))^2$$

Dodatkowo, pokaż:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n},$$

Wywnioskuj, że:

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$
$$\beta_1 = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

1. Zbiór Cars93 z biblioteki MASS - analiza regresji

Pierwszą próbką wybraną do analizy regresji jest zbiór `Cars93` z biblioteki `MASS`. Zbiór ten zawiera dane dotyczące 93 samochodów osobowych sprzedawanych w USA w 1993 roku. Zawiera on różne cechy samochodów, takie jak cena, moc silnika, liczba miejsc, itp. W tej analizie skoncentruję się na regresji liniowej, aby zbadać zależność między ceną samochodu a jego mocą silnika.

Łaďadowanie danych, stworzenie modelu regresji:

```
library(MASS)
data(Cars93)
model_cars93 <- lm(Price ~ Horsepower, data = Cars93)
```

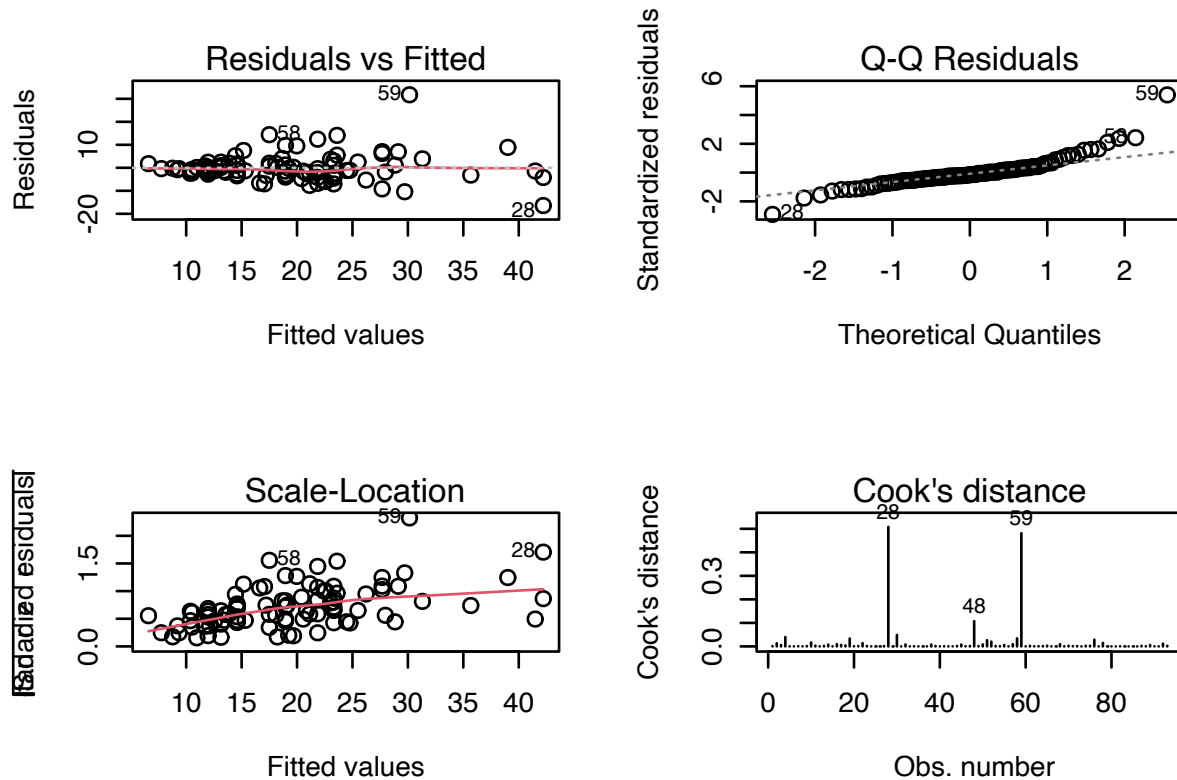
Podsumowanie modelu:

```
summary(model_cars93)

##
## Call:
## lm(formula = Price ~ Horsepower, data = Cars93)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.413  -2.792  -0.821   1.803  31.753
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.3988     1.8200  -0.769   0.444
## Horsepower    0.1454     0.0119  12.218 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.977 on 91 degrees of freedom
## Multiple R-squared:  0.6213, Adjusted R-squared:  0.6171
## F-statistic: 149.3 on 1 and 91 DF,  p-value: < 2.2e-16
```

Diagnostyka modelu regresji:

```
par(mfrow = c(2, 2))
plot(model_cars93, which = 1:4)
```



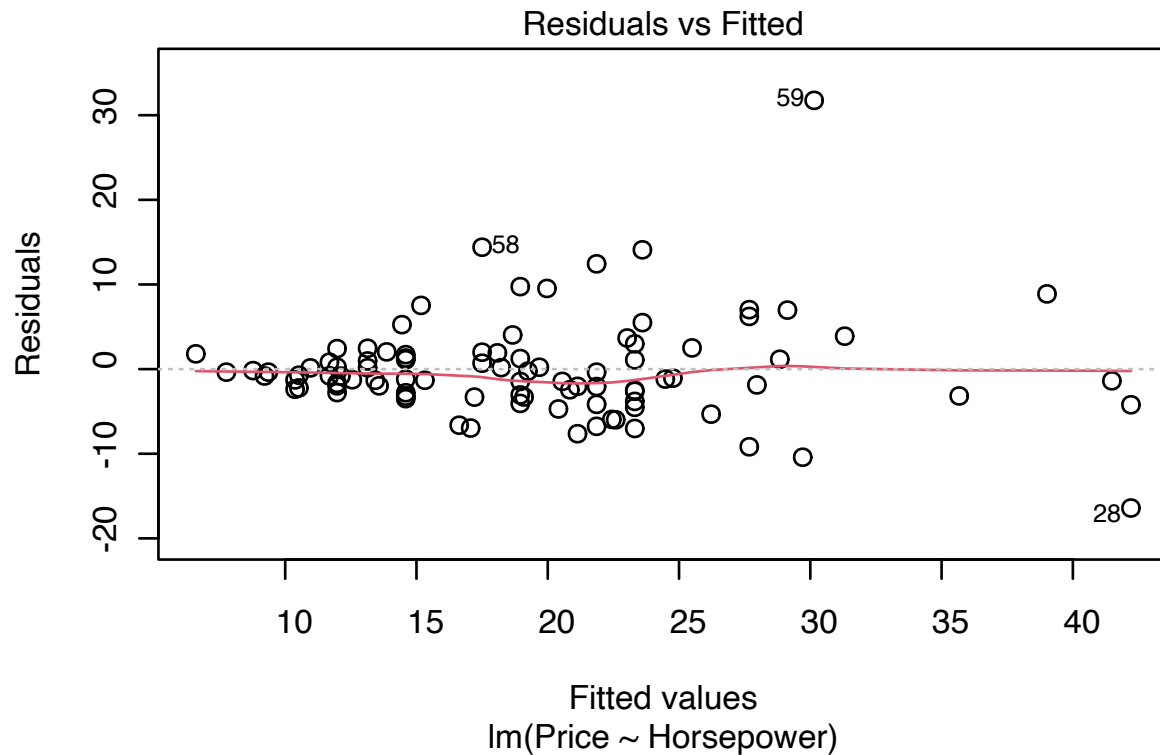
Przeznaczenie wykresów: Model regresji estymuje zależność ceny samochodu od liczby koni mechanicznych. Wykresy diagnostyczne pomagają ocenić spełnienie założeń modelu:

- **Reszty vs wartość dopasowana** – pozwala ocenić liniowość i jednorodność wariancji (homoskedastyczność).
- **Normal $Q - Q$** – ocenia, czy reszty są normalnie rozłożone (zgodne z rozkładem normalnym).
- **Scale-Location** – sprawdza jednorodność wariancji (bardziej czule niż pierwszy wykres).
- **Residuals vs Leverage** – pozwala wykryć punkty wpływowe (mające silny wpływ na model).

Przyglądnijmy się teraz każdemu z wykresów z osobna.

1.1. Wykres 1: Residuals vs Fitted

```
plot(model_cars93, which = 1)
```



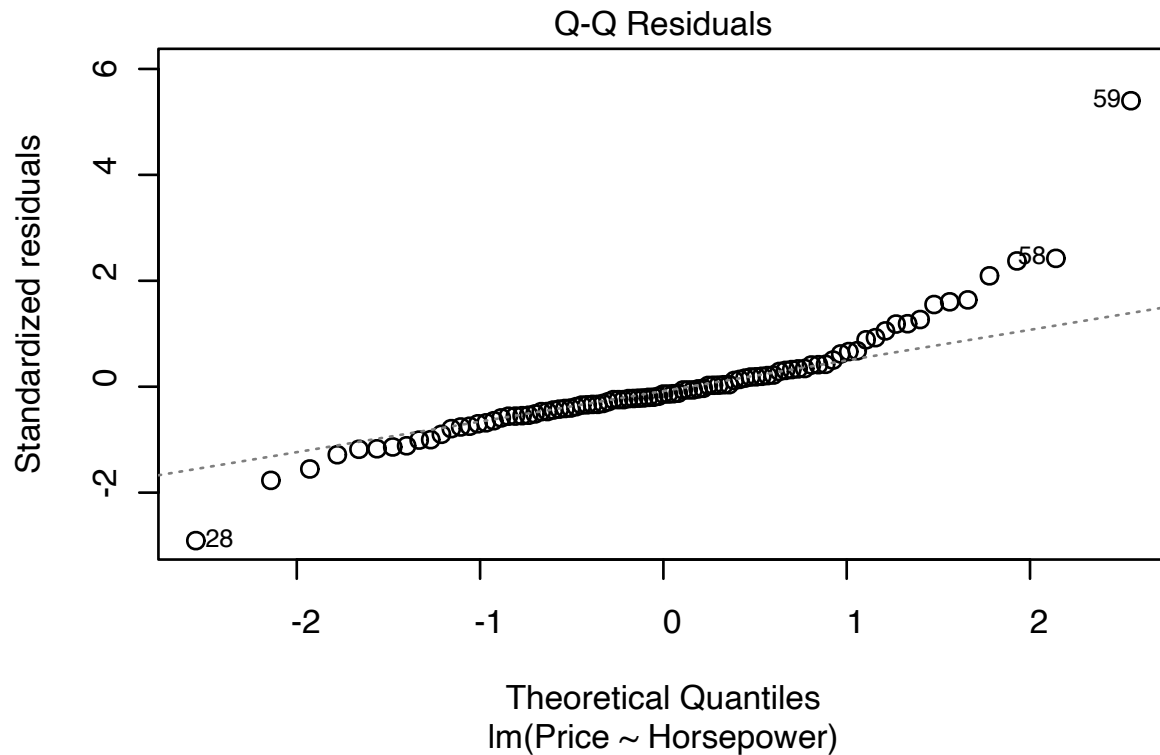
Cel: Ocena liniowości i jednorodności wariancji (homoskedastyczność).

Obserwacje: Wykres pokazuje wzrost rozrzutu reszt wraz ze wzrostem wartości dopasowanych. Idealnie, punkty powinny być rozłożone losowo wokół linii poziomej (0), co sugeruje, że model dobrze dopasowuje się do danych. W tym przypadku widać pewne odchylenia, co sugeruje **heteroskedastyczność**, czyli niejednorodność wariancji - oznacza to zmienność błędu zależną od wartości zmiennej objaśniającej. Większość punktów znajduje się w pobliżu zera, co sugeruje, że model dobrze dopasowuje się do danych. Jednakże, niektóre punkty są znacznie oddalone od linii poziomej, co może sugerować obecność punktów wpływowych lub obserwacji odstających - takich jak punkty 28, 58 i 59.

Wnioski: Naruszenie założenia homoskedastyczności może prowadzić do błędnych wniosków dotyczących istotności statystycznej współczynników regresji. W takim przypadku można rozważyć transformację zmiennej zależnej lub zastosowanie modeli regresji, które są bardziej odporne na heteroskedastyczność.

1.2. Wykres 2: Normal Q – Q Plot

```
plot(model_cars93, which = 2)
```



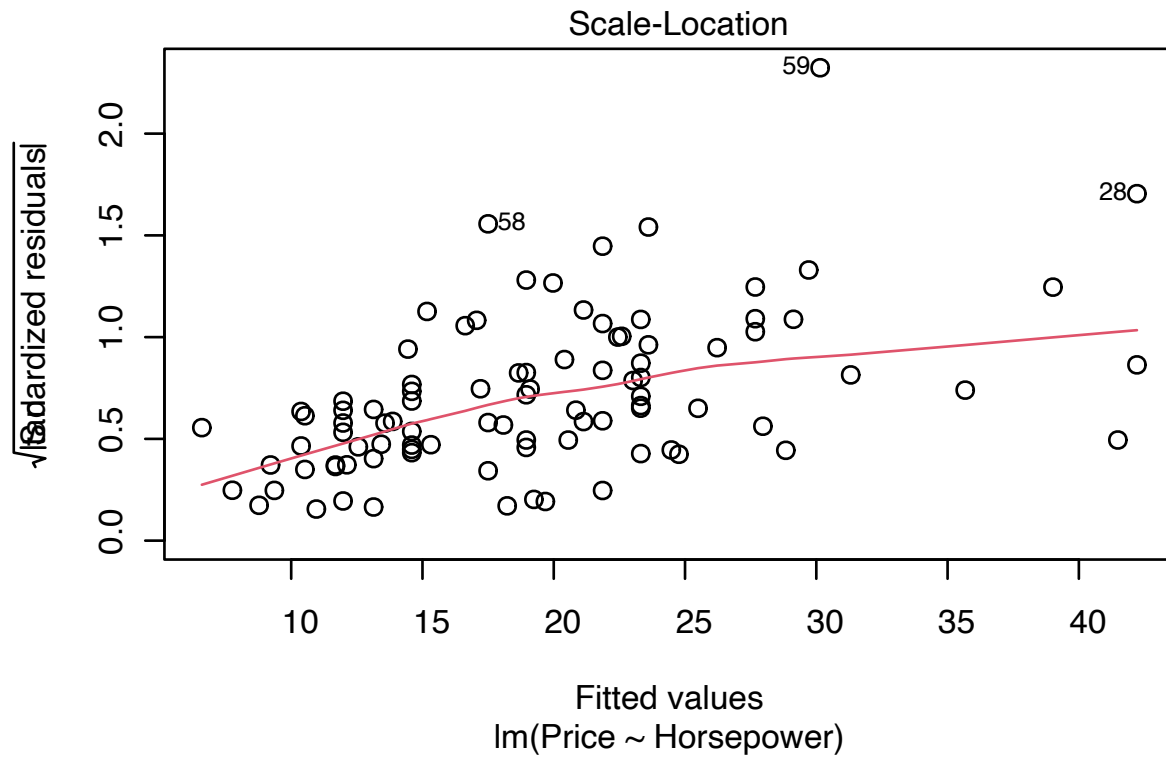
Cel: Ocena zgodności rozkładu reszt z rozkładem normalnym.

Obserwacje: W środkowej części wykresu punkty są blisko linii prostej, co sugeruje, że reszty są w miarę normalnie rozłożone. Jednakże, w końcowych częściach wykresu (lewy i prawy ogon) punkty zaczynają odbiegać od linii prostej, co sugeruje, że rozkład reszt nie jest idealnie normalny. W szczególności, punkt 59 jest istotnie wyższy niż sugerowałby rozkład normalny.

Wnioski: Reszty nie są idealnie normalnie rozłożone, co może wpływać na wyniki analizy regresji - mogą występować obserwacje odstające (outliers), które zaburzają rozkład.

1.3. Wykres 3: Scale - Location

```
plot(model_cars93, which = 3)
```



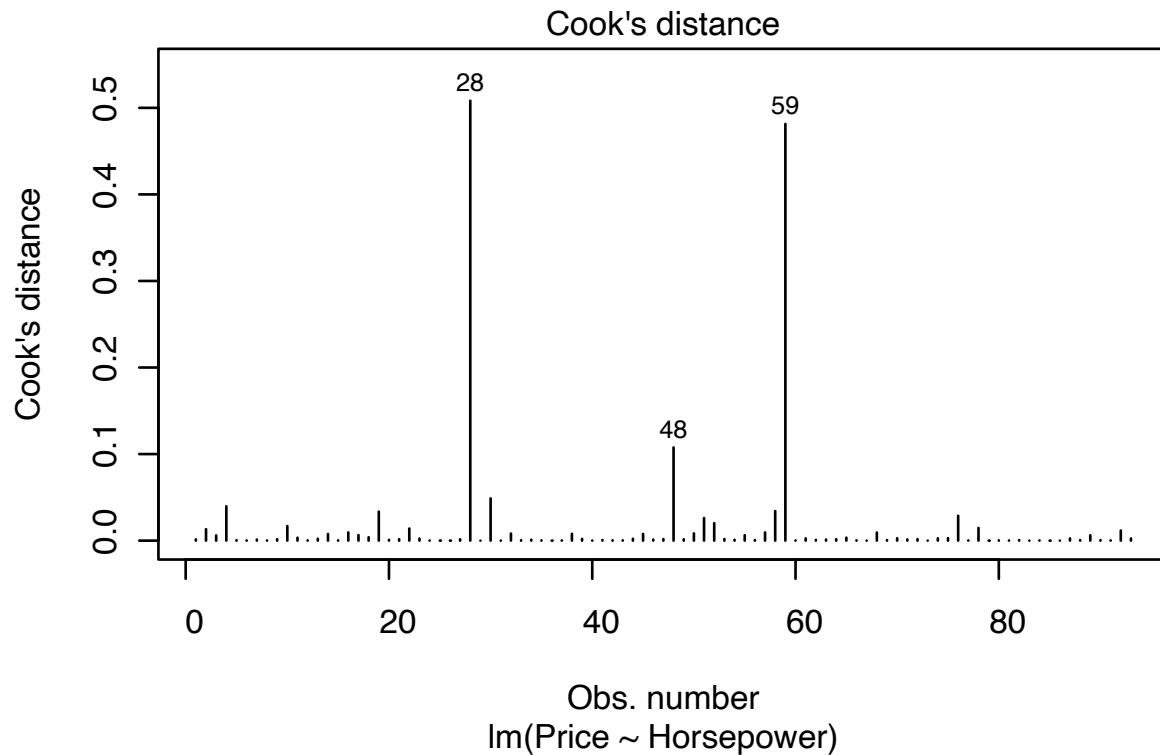
Cel: Bardziej czule sprawdzenie homoskedastyczności

Obserwacje: Wraz z wartościami dopasowanymi rosną wartości standaryzowanych reszt, co jest klasycznym objawem heteroskedastyczności - wariancja rośnie.

Wnioski: Potwierdza to poprzednie przypuszczenie - model narusza założenie stałej wariancji reszt.

1.4. Wykres 4: Cook's Distance

```
plot(model_cars93, which = 4)
```



Cel: Zidentyfikowanie obserwacji mających silny wpływ na model.

Obserwacje: Punkty 28, 48 i 59 mają najwyższe wartości odległości Cooka - są to obserwacje wpływowe. Szczególnie punkty 28 i 59 - mogą one istotnie zmieniać wynik estymacji, jeśli zostaną usunięte.

Wnioski: Warto rozważyć ponowną analizę modelu z pominięciem tych obserwacji, lub zastosowanie metod odpornych, takich jak np. regresja odporna (robust regression).

Podsumowanie

Table 1: Podsumowanie diagnostyki modelu regresji Cars93

Kwestia	Ocena	Komentarz
Liniowość	Częściowo spełniona	Brak wyraźnych wzorców nieliniowych
Homoskedastyczność	Naruszona	Widoczne zwiększenie wariancji błędu
Normalność reszt	Naruszona (na końcach)	Wskazuje na odstające wartości
Obserwacje wpływowe	Zidentyfikowane	28, 48, 59 – warto monitorować

1.5. Wyznaczenie współczynników regresji - metoda najmniejszych kwadratów

Rozważamy model liniowy:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i,$$

gdzie:

- Y_i to cena samochodu (Price),
- X_i to liczba koni mechanicznych (Horsepower).

Teraz wyznaczmy współczynniki regresji b_0 i b_1 minimalizując sumę kwadratów reszt. Wykorzystajmy do tego następujący wzór:

$$\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - (b_0 + b_1 X_i))^2.$$

```
# Dane x i y
x <- Cars93$Horsepower
y <- Cars93$Price
n <- length(x)
# Krok 1: Średnie
x_bar <- mean(x)
y_bar <- mean(y)
# Krok 2: Składniki wzorów analitycznych
suma_x2 <- sum(x^2)
suma_x <- sum(x)
suma_y <- sum(y)
suma_xy <- sum(x * y)
# Pokazujemy wzór wariancji z dwóch składników:
suma_kwadratow_x <- suma_x2 - (suma_x^2) / n
# Współczynnik kowariancji:
suma_kowariancji <- suma_xy - (suma_x * suma_y) / n
# Wyznaczamy współczynnik nachylenia (b1)
b1 <- suma_kowariancji / suma_kwadratow_x
# Wyznaczamy wyraz wolny (b0)
b0 <- y_bar - b1 * x_bar
```

Wyniki:

```
## sum((x_i - x̄)^2) = 252363.2
```

```
## sum((x_i - x̄)(y_i - ȳ)) = 36686.35
```

Table 2: Współczynniki regresji Cars93

Wspolczynnik	Wartosc
b0 (nachylenie)	-1.3987691
b1 (wyraz wolny)	0.1453712

```
## Interpretacja: Przeciętnie, cena samochodu rośnie o ok. 145.4 dolarów za każdy
## dodatkowy koń mechaniczny.
```

1.6. Dowód algebraiczny - rozkład wariancji (suma kwadratów)

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n},$$

```
lhs_var <- sum((x - mean(x))^2)
rhs_var <- sum(x^2) - (sum(x)^2) / n

## sum((x_i - x̄)^2) = 252363.2
## sum(x_i^2) - (sum(x_i))^2/n = 252363.2
all.equal(lhs_var, rhs_var)

## [1] TRUE
```

Wynik: Obie strony równania są równe, co potwierdza poprawność wzoru.

1.7. Wyznaczenie wzoru na współczynnik regresji β_1

Rozważmy klasyczny model regresji liniowej:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

Celem jest wyprowadzenie estymatora nachylenia β_1 , który minimalizuje sumę kwadratów reszt:

$$\min \sum (Y_i - (b_0 + b_1 X_i))^2$$

Z rachunku różniczkowego wiadomo, że minimalizując sumę kwadratów, otrzymujemy estymator:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Z własności algebry sum można rozwinąć licznik tego wzoru:

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

Analogicznie, mianownik można zapisać jako:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

Z definicji kowariancji i wariancji:

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Podstawiając je do wzoru na β_1 :

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{n \cdot \text{cov}(x, y)}{n \cdot \text{var}(x)} = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

Wynik: Otrzymaliśmy wzór na współczynnik regresji β_1 jako stosunek kowariancji do wariancji.

$$\beta_1 = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

czyli nachylenie prostej regresji jest równe stosunkowi kowariancji zmiennych do wariancji zmiennej niezależnej.

2. Zbiór USCrime z biblioteki MASS - analiza regresji

Drugą próbkę wybraną do analizy regresji stanowi zbiór `USCrime` z biblioteki `MASS`. Dane te dotyczą przestępczości w 47 stanach USA i zawierają wiele zmiennych społeczno-ekonomicznych, takich jak nierówność dochodów, ilość przestępstw na osobę, wydatki na policję czy też średnią liczbę lat nauki. W tej analizie skoncentruję się na badaniu wpływu wskaźnika nierówności dochodów (`Ineq`) na przestępczość na mieszkańca (`y`).

Załadowanie danych i stworzenie modelu regresji:

```
library(MASS)
data(USCrime)
model_uscrime <- lm(y ~ Ineq, data = USCrime)
```

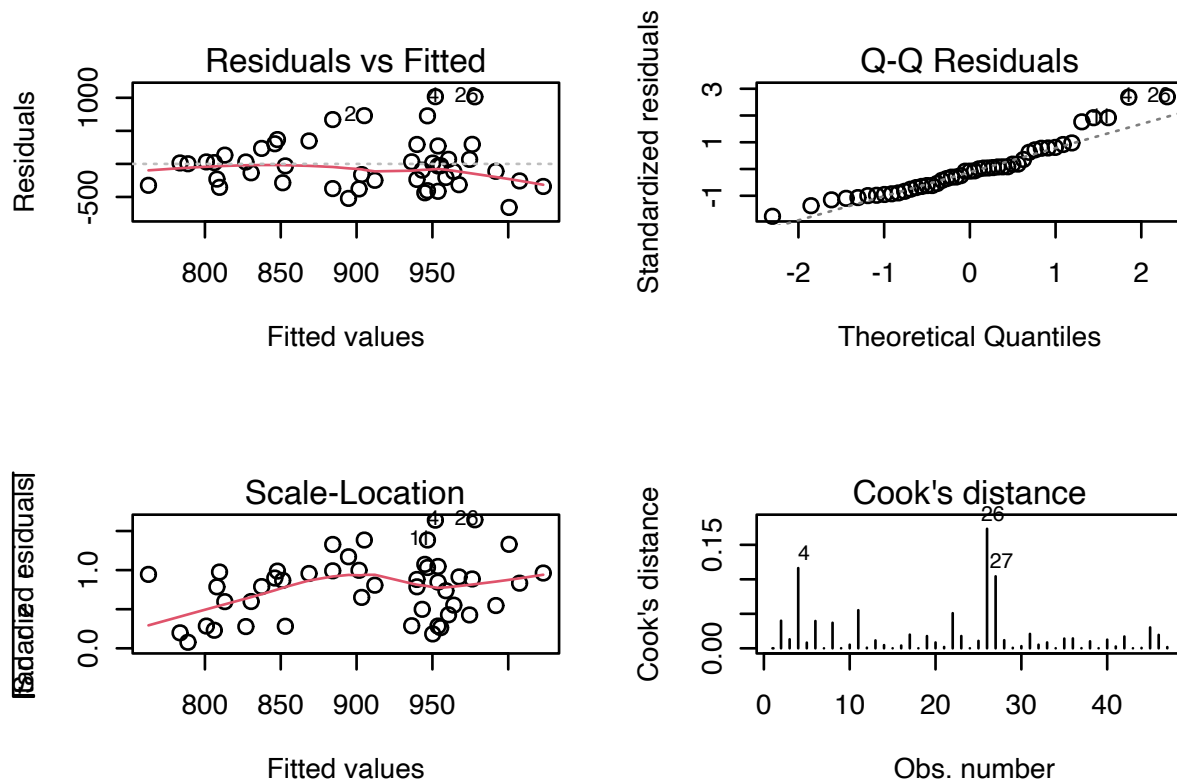
Podsumowanie modelu:

```
summary(model_uscrime)

##
## Call:
## lm(formula = y ~ Ineq, data = USCrime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -658.54 -271.38  -30.02  183.75 1017.06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1241.773    281.478   4.412 6.33e-05 ***
## Ineq         -1.736      1.422  -1.221   0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 384.7 on 45 degrees of freedom
## Multiple R-squared:  0.03205,    Adjusted R-squared:  0.01054
## F-statistic: 1.49 on 1 and 45 DF,  p-value: 0.2286
```

Diagnostyka modelu regresji:

```
par(mfrow = c(2, 2))
plot(model_uscrime, which = 1:4)
```



Analiza wykresów: Model regresji estymuje zależność liczby przestępstw od wskaźnika nierówności dochodów w danym stanie. Wykresy diagnostyczne umożliwiają ocenę spełnienia kluczowych założeń regresji.

Residuals vs Fitted - wykres pokazuje wartości reszt w miarę losowo rozłożone wokół zera, co sugeruje, że model dobrze dopasowuje się do danych. Nie widać wyraźnych wzorców nieliniowych, choć niektóre punkty są bardziej oddalone od linii poziomej - 2 oraz 26, co może sugerować obecność punktów wpływowych, a w konsekwencji nieliniowość czy też heteroskedastyczność.

Normal Q - Q - punkty w pewien sposób odbiegają od linii, zwłaszcza na końcach rozkładu - punkt 26. Sugeruje to, że dla tych danych nie można przyjąć założenia o normalności rozkładu.

Scale-Location - widzimy tutaj niewielką tendencję do danych wzrostowych, co sugeruje niejednorodność wariancji - wartości reszt mogą rosnąć wraz z wartością dopasowaną.

Cook's Distance - punkty 4, 26 oraz 27 odbiegają od reszty, co sugeruje, że mogą mieć one silny wpływ na model. Warto je monitorować i rozważyć ich usunięcie z analizy.

Podsumowanie

Table 3: Podsumowanie diagnostyki modelu regresji UScrime

Kwestia	Ocena	Komentarz
Liniowość	Możliwe naruszenie	Lekkie zakrzywienie w trendzie liniowym
Homoskedastyczność	Naruszona	Widoczne zwiększenie wariancji błędu
Normalność reszt	Naruszona	Wskazuje na odstające wartości, zwłaszcza na końcach
Obserwacje wpływowe	Zidentyfikowane	4, 26, 27- warto monitorować

2.1. Wyznaczenie współczynników regresji - metoda najmniejszych kwadratów

Rozważamy model liniowy:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i,$$

gdzie:

- Y_i to wskaźnik przestępczości (y),
- X_i to wskaźnik nierówności dochodów w stanie (Ineq).

Teraz wyznaczmy współczynniki regresji b_0 i b_1 minimalizując sumę kwadratów reszt. Wykorzystajmy do tego następujący wzór:

$$\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - (b_0 + b_1 X_i))^2.$$

```
# Dane x i y
x <- UScrime$Ineq
y <- UScrime$y
n <- length(x)
# Krok 1: Średnie
x_bar <- mean(x)
y_bar <- mean(y)
# Krok 2: Składniki wzorów analitycznych
suma_x2 <- sum(x^2)
suma_x <- sum(x)
suma_y <- sum(y)
suma_xy <- sum(x * y)
# Pokazujemy wzór wariancji z dwóch składników:
suma_kwadratow_x <- suma_x2 - (suma_x^2) / n
# Współczynnik kowariancji:
suma_kowariancji <- suma_xy - (suma_x * suma_y) / n
# Wyznaczamy współczynnik nachylenia (b1)
b1 <- suma_kowariancji / suma_kwadratow_x
# Wyznaczamy wyraz wolny (b0)
b0 <- y_bar - b1 * x_bar
```

Wyniki:

```
## sum((x_i - x̄)^2) = 73218
```

```
## sum((x_i - x̄)(y_i - ȳ)) = -127070
```

Table 4: Współczynniki regresji USCrime ($y \sim \text{Ineq}$)

Wspolczynnik	Wartosc
b0 (wyraz wolny)	1241.7725
b1 (nachylenie)	-1.7355

```
## Interpretacja: Przeciętnie, przy wzroście wskaźnika nierówności dochodów o 1 punkt,
## wskaźnik przestępczości rośnie o ok. 1.74 jednostek.
```

. Dowód algebraiczny - rozkład wariancji (suma kwadratów)

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n},$$

```
lhs_var <- sum((x - mean(x))^2)
rhs_var <- sum(x^2) - (sum(x)^2) / n
```

```
## sum((x_i - x̄)^2) = 73218
```

```
## sum(x_i^2) - (sum(x_i))^2/n = 73218
```

```
all.equal(lhs_var, rhs_var)
```

```
## [1] TRUE
```

Wynik: Obie strony równania są równe, co potwierdza poprawność wzoru.

3. Zbiór road z biblioteki MASS - analiza regresji

Jako ostatni dataset wybrałem zbiór `road` z biblioteki `MASS`. Dane te przedstawiają liczbę ofiar śmiertelnych w wypadkach drogowych w 25 stanach USA, a także liczbę kierowców, gęstość zaludnienia, długość dróg wiejskich, temperatury, zużycie paliwa oraz inne czynniki mogące mieć wpływ na wypadkowość. W tej analizie skupimy się na zbadaniu zależności między liczbą kierowców (`drivers`) a liczbą zgonów w wypadkach drogowych (`deaths`).

Łaďadowanie danych i stworzenie modelu regresji:

```
library(MASS)
data(road)
model_road <- lm(deaths ~ drivers, data = road)
```

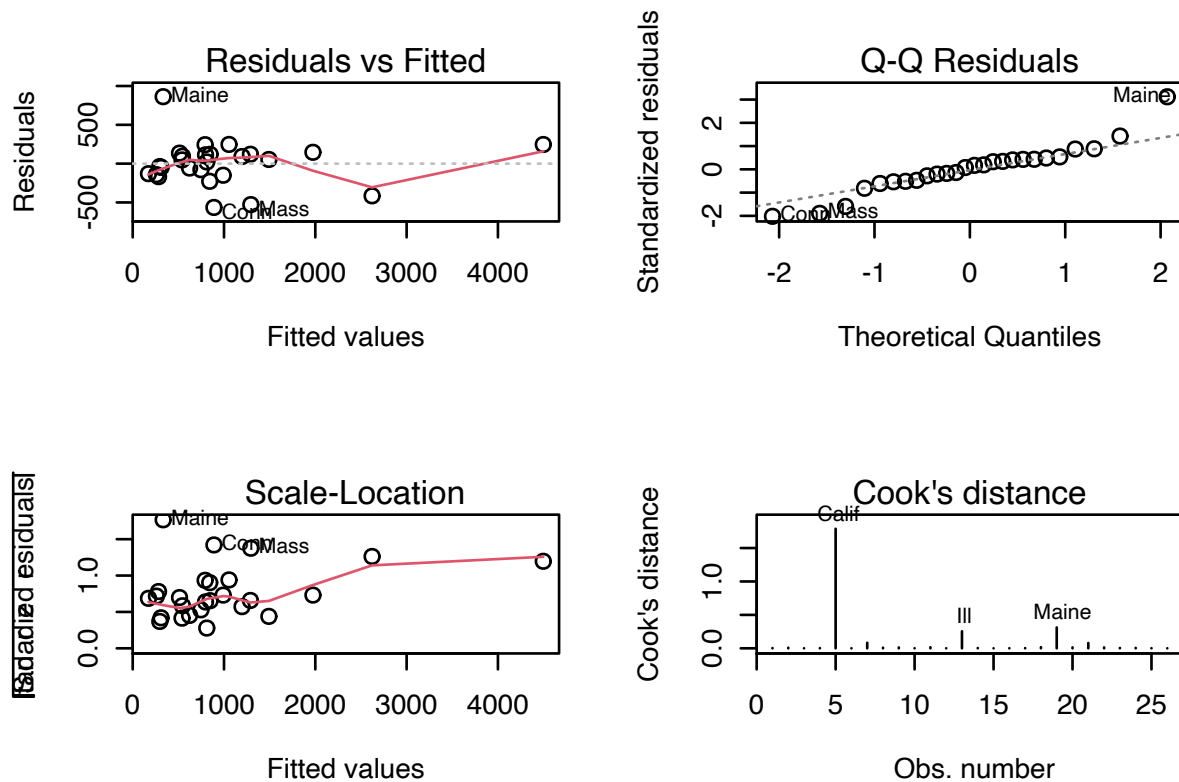
Podsumowanie modelu:

```
summary(model_road)

##
## Call:
## lm(formula = deaths ~ drivers, data = road)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -564.49 -138.88   34.69  120.52  862.52
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 122.0989    78.7076   1.551   0.134
## drivers      4.5951     0.2897  15.863 3.19e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 285.2 on 24 degrees of freedom
## Multiple R-squared:  0.9129, Adjusted R-squared:  0.9093
## F-statistic: 251.6 on 1 and 24 DF,  p-value: 3.192e-14
```

Diagnostyka modelu regresji:

```
par(mfrow = c(2, 2))
plot(model_road, which = 1:4)
```

Analiza wykresów: Model regresji estymuje zależność liczby kierowców od liczby zgonów w wypadkach drogowych. Wykresy diagnostyczne umożliwiają ocenę spełnienia kluczowych założeń regresji.

Residuals vs Fitted - na wykresie obserwujemy, że punkty nie są rozłożone losowo wokół poziomej linii zera. Widać pewien wzór, sugerujący potencjalną nieliniowość w danych. Dodatkowo, wydaje się, że rozrzut reszt zwiększa się wraz ze wzrostem wartości dopasowanych, co może wskazywać na heteroskedastyczność. Punkty *Maine* oraz *Mass* wydają się znacznie odbiegać od reszty.

Normal Q-Q - punkty w pewien sposób odbiegają od linii, zwłaszcza na końcach rozkładu - punkt *Maine*. Sugeruje to, że dla tych danych nie można przyjąć założenia o normalności rozkładu.

Scale-Location - widzimy tutaj wzrost pierwiastka kwadratu standaryzowanych reszt wraz ze wzrostem wartości dopasowanych, co sugeruje niejednorodność wariancji, ponadto punkty *Maine* oraz *Mass* wydają się mieć ponownie wyższe wartości na tym wykresie.

Cook's Distance - punkty *Maine*, *Ill* oraz *Calif* wydają się mieć szczególnie duży wpływ na model. Warto je monitorować i rozważyć ich usunięcie z analizy.

Podsumowanie wykresów diagnostycznych

Table 5: Podsumowanie diagnostyki modelu regresji road

Kwestia	Ocena	Komentarz
Liniowość	Naruszona	Widoczny wzór nieliniowy i zmiana wariancji
Homoskedastyczność	Naruszona	Rozrzut reszt wraz ze wzrostem wartości dopasowanych
Normalność reszt	Naruszona	Odstępstwa od normalności, Maine oraz Mass znacznie odstają
Obserwacje wpływowe	Zidentyfikowane	Maine, Ill oraz Calif wykazują szczególnie duży wpływ

3.1. Wyznaczenie współczynników regresji - metoda najmniejszych kwadratów

Rozważamy model liniowy:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i,$$

gdzie:

- Y_i to liczba zgonów w wypadkach drogowych (`deaths`),
- X_i to liczba kierowców w stanie (`drivers` - w dziesiątkach tysięcy).

Teraz wyznaczmy współczynniki regresji b_0 i b_1 minimalizując sumę kwadratów reszt. Wykorzystajmy do tego następujący wzór:

$$\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - (b_0 + b_1 X_i))^2.$$

```
# Dane x i y
x <- road$drivers
y <- road$deaths
n <- length(x)
# Krok 1: Średnie
x_bar <- mean(x)
y_bar <- mean(y)
# Krok 2: Składniki wzorów analitycznych
suma_x2 <- sum(x^2)
suma_x <- sum(x)
suma_y <- sum(y)
suma_xy <- sum(x * y)
# Pokazujemy wzór wariancji z dwóch składników:
suma_kwadratow_x <- suma_x2 - (suma_x^2) / n
# Współczynnik kowariancji:
suma_kowariancji <- suma_xy - (suma_x * suma_y) / n
# Wyznaczamy współczynnik nachylenia (b1)
b1 <- suma_kowariancji / suma_kwadratow_x
# Wyznaczamy wyraz wolny (b0)
b0 <- y_bar - b1 * x_bar
```

Wyniki:

```
## sum((x_i - x̄)^2) = 969024
```

```
## sum((x_i - x̄)(y_i - ȳ)) = 4452799
```

Table 6: Współczynniki regresji road (`deaths ~ drivers`)

Współczynnik	Wartość
b0 (wyraz wolny)	122.0989
b1 (nachylenie)	4.5951

```
## Interpretacja: Przeciętnie, liczba zgonów w wypadkach drogowych wzrasta o około 4.6 na każde
## dodatkowe 10 000 kierowców w stanie.
```

3.2. Dowód algebraiczny - rozkład wariancji (suma kwadratów)

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n},$$

```
lhs_var <- sum((x - mean(x))^2)
rhs_var <- sum(x^2) - (sum(x)^2) / n

## sum((x_i - x̄)^2) = 969024
## sum(x_i^2) - (sum(x_i))^2/n = 969024
all.equal(lhs_var, rhs_var)

## [1] TRUE
```

Wynik: Obie strony równania są równe, co potwierdza poprawność wzoru.