

Testy statystyczne

Karolina Popiołek

2024-08-09

1. Dane

W niniejszym raporcie będę pracować na zbiorze danych "mtcars", który obejmuje informacje odnośnie poszczególnych modeli samochodów. Próba składa się z 32 obserwacji. Poniżej znajduje się podgląd zbioru.

```
data(mtcars)
head(mtcars)
```

##		mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
##	Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
##	Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
##	Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
##	Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
##	Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
##	Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

2. Badanie zależności zużycia paliwa (mpg) od masy samochodu (wt) i mocy silnika (hp)

$$\text{mpg} = B_0 + B_1 * \text{wt} + B_2 * \text{hp} + \epsilon$$

Założenia klasycznej regresji liniowej:

1. Normalność rozkładu składnika losowego
2. Liniowość
3. Homoskedastyczność (stałość wariancji składnika losowego)
4. Brak autokorelacji: Składnik losowy

```
model <- lm(mpg ~ wt + hp, data = mtcars)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ wt + hp, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727    1.59879   23.285 < 2e-16 ***
## wt          -3.87783    0.63273   -6.129 1.12e-06 ***
## hp           -0.03177    0.00903   -3.519 0.00145 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.593 on 29 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
## F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

Wnioski:

1. Wzrost masy samochodu o jedną jednostkę prowadzi do spadku zużycia paliwa średnio o 3.88, przy założeniu stałości mocy silnika.
2. Wzrost mocy silnika o jedną jednostkę prowadzi do spadku zużycia paliwa średnio o 0.032, przy założeniu stałości masy pojazdu.
3. Obie zmienne mają znaczący wpływ na zmienną mpg, a zatem zmiana masy samochodu i zmiana mocy silnika są silnie związane z zużyciem paliwa.
4. Wysoka wartość statystyki F i bardzo niskie p-value wskazują na to, że model jest istotny statystycznie.

2. Testy statystyczne

2.1. Testy normalności Shapiro-Wilka i Jarque-Bera

Sprawdzam, czy próbka pochodzi z rozkładu normalnego. Przy obu testach stawiam następujące hipotezy:

H0: Rozkład jest normalny H1: Rozkład nie jest normalny

```
shapiro.test(residuals(model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(model)
## W = 0.92792, p-value = 0.03427
```

```
jarque.bera.test(residuals(model))
```

```
##
##  Jarque Bera Test
##
## data:  residuals(model)
## X-squared = 4.0462, df = 2, p-value = 0.1322
```

Interpretacja Shapiro-Wilka: $p\text{-value} < 0,05$, a zatem na poziomie istotności $0,05$ odrzucamy H_0 i zakładamy, że model nie ma rozkładu normalnego.

Interpretacja Jarque-Bera: $p\text{-value} > 0,05$, a zatem nie ma podstaw do odrzucania H_0 i zakładamy, że nie występują istotne odchylenia od normalności.

Rozbieżność wynika z faktu, iż test Shapiro-Wilka jest bardziej czuły niż test Jarque-Bera, zwłaszcza przy małych próbach. Decyduję się kontynuować analizę, mając na uwadze, że rozkład nie jest idealnie normalny.

2.2. Test homoskedastyczności - Test Breuscha-Pagana

W celu sprawdzenia homoskedastyczności stawiam następujące hipotezy: H_0 : Wariancja reszt jest stała H_1 : Wariancja reszt nie jest stała

```
bptest(model)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: model  
## BP = 0.88072, df = 2, p-value = 0.6438
```

Interpretacja: $p\text{-value} > 0.05$, a zatem nie ma podstaw do odrzucenia hipotezy zerowej o braku heteroskedastyczności (zakładamy, że wariancja jest stała).

2.3. Test autokorelacji Durbina-Watsona

Założenia: H_0 : Autokorekacja reszt nie występuje H_1 : Występuje autokorelacja reszt

```
dwtest(model)
```

```
##  
## Durbin-Watson test  
##  
## data: model  
## DW = 1.3624, p-value = 0.02061  
## alternative hypothesis: true autocorrelation is greater than 0
```

Interpretacja: Ponieważ statystyka $D-W < 2$ istnieje możliwość występowania dodatniej autokorelacji. Z uwagi na wartość $p\text{-value}$, która jest mniejsza niż poziom ufności $0,05$, mamy podstawy do odrzucenia hipotezy zerowej i przyjęcia, że istnieje dodatnia autokorelacja, a więc reszty mogą być ze sobą skorelowane. W związku z tym nie zostało spełnione jedno z założeń regresji liniowej, więc model odrzucamy jako nieprawidłowy.