

W pewnej klinice badano związek między aktywnością enzymów aminotransferazy a stężeniem amoniaku we krwi u chorych z ostrą niewydolnością wątroby. Pobrano losową próbę 10 pacjentów i otrzymano następujące wyniki:

aktywność	430	470	520	570	630	690	740	770	800	780
stężenie	31	33	36	39	42	47	51	54	55	57

1. Wizualna ocena zależności (za pomocą pakietu R)

Wprowadzenie danych

```
> aktywność <- c(430,470,520,570,630,690,740,770,800,780)
```

```
> stężenie <- c(31,33,36,39,42,47,51,54,55,57)
```

Utworzenie wykresu

```
> plot(aktywność,stężenie,pch=19,col="red")
```

Punkty na wykresie są wypełnione: pch=19.

Punkty są koloru czerwonego: col="red".

Ocena wykresu daje pewne podstawy do przyjęcia następującego modelu:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

gdzie ε_i , $i = 1, 2, \dots, n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie $N(0, \sigma^2)$.

Uwaga: Model dotyczy rozkładu warunkowego $Y|X = x$, ponieważ obie badane cechy są cechami losowymi.

2. Weryfikacja hipotezy o braku zależności między cechami.

Brak zależności oznacza, że poziomu amoniaku we krwi nie da się określać za pomocą poziomu aktywności enzymów. Przy założeniu, że zaproponowany model liniowy jest poprawny, hipotezę o braku zależności zapisujemy w postaci:

$$H_0 : \beta_1 = 0$$

Do weryfikacji tej hipotezy, na zadanym poziomie istotności $\alpha = 0.05$, zastosujemy test F .

Implementacja modelu

```
> model <- lm(stężenie ~ aktywność)
```

Podstawowe wyniki

```
> wyniki <- summary(model)
```

```
> wyniki
```

Ostatni wiersz wyników: F-statistic: 452.5 on 1 and 8 DF, p-value: 2.509e-08

Ponieważ p-value < 0.05, hipotezę odrzucamy.

Wniosek: Jest zależność między cechami.

3. Opis ilościowy zależności.

Wyświetlenie współczynników

```
> model
```

```
(Intercept) aktywność
```

```
-0.18503 0.06982
```

Oszacowana funkcja regresji: $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = -0.18503 + 0.06982x$

Interpretacja współczynnika kierunkowego funkcji regresji: Współczynnik kierunkowy pokazuje przeciętną różnicę między pacjentami w badanej populacji. Jest to przeciętna różnica w stężeniu amoniaku we krwi, jaka przypada na jednostkową różnicę w aktywności enzymów. Oszacowanie punktowe tej różnicy wynosi

0.06982. O precyzji tego oszacowania świadczy szerokość przedziału ufności. Przedział ufności dla parametru β_1 , na poziomie ufności 0.95, wynosi (0.06225104, 0.07738968). Obrazowo można powiedzieć, że pacjenci, których aktywności enzymów różnią się o 10, to ze względu na stężenie amoniaku we krwi, różnią się średnio o wartość z przedziału (0.6225104, 0.7738968) \approx (0.62, 0.77).

Przedziały ufności dla parametrów funkcji regresji (na poziomie ufności $1 - \alpha = 0.95$).

```
> confint(model,level=0.95)
(Intercept) -5.12716355 4.75710367
aktywność 0.06225104 0.07738968
```

4. Przedział ufności dla funkcji regresji oraz przedział predykcji.

Problem 1. Ile przeciętnie wynosi poziom amoniaku we krwi osób o poziomie aktywności enzymów równym 80?

```
> predict(model,data.frame(aktywność=c(80)),level=0.95,interval="confidence")
      fit      lwr      upr
1  5.400599  1.05038  9.750818
```

Wynosi on co najmniej 1.05038, ale nie więcej niż 9.750818.

Problem 2. Jaka jest rozpiętość stężenia amoniaku we krwi osób, których aktywność enzymów wynosi 80

```
> predict(model,data.frame(aktywność=c(80)),level=0.95,interval="prediction")
      fit      lwr      upr
1  5.400599  0.06275274 10.73844
```

Stężenie amoniaku wynosi co najmniej 0.06275274, ale nie więcej niż 10.73844 wśród 95% osób o aktywności enzymów na poziomie 80.

5. Weryfikacja poprawności modelu

Powyższa analiza została przeprowadzona przy założeniu, że zaproponowany model liniowy dobrze opisuje dane rzeczywiste. Jeżeli je źle opisuje, analiza ta nie ma sensu.

Wykres kwantylowy	Wykresy reszt	stężenie vs oszacowane stężenie
<code>> qqnorm(model\$res)</code>	<code>> plot(model\$fitted.values,model\$residuals)</code>	<code>> plot(model\$fitted.values,stężenie)</code>
<code>> qqline(model\$res)</code>	<code>> abline(h=0)</code>	<code>> abline(0,1)</code>

Z wykresów wynika, że model może być źle dopasowany. Ewentualna zamiana zależności liniowej na krzywo-liniową wykaże obserwacje odstające. Obserwacje te należy zweryfikować. Jeżeli weryfikacja nie da podstaw do ich odrzucenia, należy zwiększyć próbę aby sprawdzić, czy zauważalna krzywoliniowość nie jest wynikiem przypadku.