

REGRESJA LINIOWA

Zacznijmy od najprostszego modelu regresji liniowej, to jest od modelu regresji liniowej jednokrotnej.

1 Regresja liniowa jednokrotna

Chcemy opisać zależność Y (tzw. *zmiennej odpowiedzi, objaśnianej lub zależnej*) od x (tzw. *zmiennej objaśniającej lub niezależnej*), wykorzystując informacje zawarte w n parach:

$$(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n),$$

gdzie x_i to i -ta wartość zmiennej objaśniającej, zaś Y_i to odpowiadająca jej wartość zmiennej objaśnianej. Zanim przeprowadzimy eksperyment i zbierzemy dane, nie znamy wartości Y_i , $i = 1, 2, \dots, n$. Zależć będą one od tego, które zdarzenie elementarne wybierzemy (np. w którym momencie będziemy zbierać dane lub które obiekty weźmiemy do badania), zatem Y_i są zmiennymi losowymi. Wartości Y_i , $i = 1, 2, \dots, n$ dla ustalonego zdarzenia losowego, oznaczać będziemy małymi literami: y_i , $i = 1, 2, \dots, n$.

Np. możemy badać zależność rekordowego czasu pokonania trasy w biegu przełajowym od długości tej trasy. Wówczas zmienną objaśnianą Y jest rekordowy czas pokonania trasy a zmienną objaśniającą x - długość pokonywanej trasy. Zebrane dane dotyczące n tras biegów przełajowych zapiszemy w postaci n par

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n),$$

gdzie x_i to długość i -tej trasy przełajowej zaś y_i to rekordowy czas pokonania tej trasy podczas biegu przełajowego.

Model jednokrotnej regresji liniowej wygląda następująco:

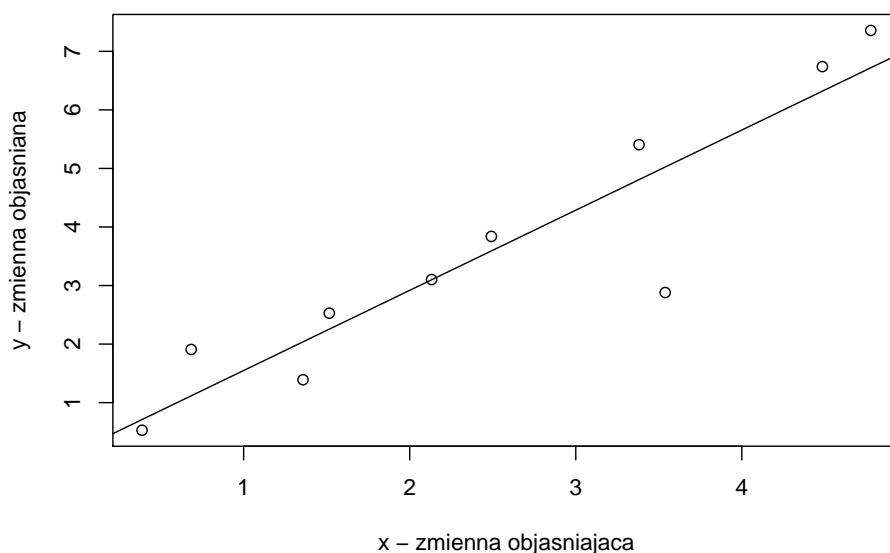
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \text{ gdzie } \varepsilon_i \text{ to niezależne zmienne losowe o rozkładzie } \mathcal{N}(0, \sigma^2),$$

natomiast β_0, β_1 i $\sigma^2 > 0$ to pewne stałe rzeczywiste (tzw. parametry modelu).

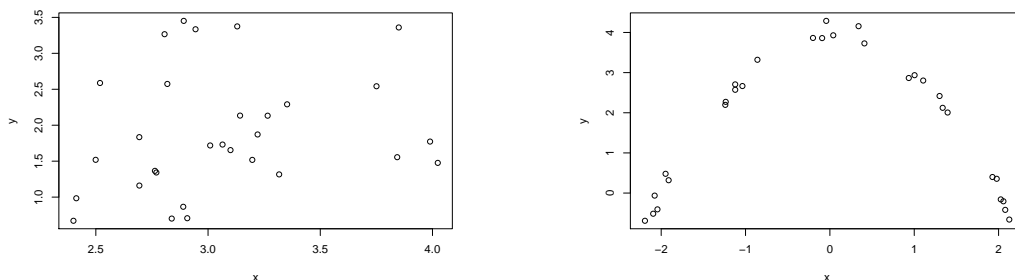
Zauważmy, że w modelu tym

$$E(Y_i) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, n,$$

co oznacza, że zależność wartości średniej zmiennej objaśnianej od zmiennej objaśniającej jest zależnością liniową. Zatem dopasowywanie modelu regresji liniowej do konkretnych danych ma sens jedynie gdy wykres rozproszenia, sporządzony dla tych danych, ma choć w przybliżeniu charakter liniowy. Przykład takiej sytuacji przedstawia poniższy rysunek:



Natomiast jeśli dysponujemy chmurą punktów nie wykazujących tendencji liniowej, to dopasowywanie prostej regresji mija się z celem. Sytuacja taka jest zobrazowana na poniższych dwóch rysunkach.



Dopasowanie do danych modelu regresji liniowej sprowadza się do wyznaczania nieznanych parametrów tego modelu β_0 , β_1 i σ^2 . Wyznaczone oszacowania parametrów β_0 , β_1 i σ^2 będziemy nazywali ich *estymatorami* i będziemy je oznaczać odpowiednio przez $\hat{\beta}_0$, $\hat{\beta}_1$ i $\hat{\sigma}^2$.

Metody wyznaczania estymatorów β_0 , β_1 i σ^2

Sposób I - metoda najmniejszych kwadratów (jedynie dla β_0 i β_1)

Metoda ta polega na dobraniu $\hat{\beta}_0$ i $\hat{\beta}_1$ w taki sposób by linia prosta $y = \hat{\beta}_0 + \hat{\beta}_1 x$ była najlepiej dopasowaną prostą do punktów $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$. Za miarę dopasowania prostej $y = b_0 + b_1 x$ do punktów $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ przyjmujemy sumę kwadratów błędów:

$$\sum_{i=1}^n \underbrace{(y_i - (b_0 + b_1 x_i))^2}_{=e_i \text{ (i-ty błąd)}}$$

- im ta suma jest mniejsza, tym dopasowanie lepsze. Zadanie sprowadza się zatem do znalezienia b_0 i b_1 , dla których funkcja dwóch zmiennych

$$f(b_0, b_1) := \sum_{i=1}^n (y_i - (b_0 + b_1 x_i))^2$$

osiąga wartość najmniejszą. Aby wyznaczyć punkt (b_0, b_1) , w którym funkcja f osiąga minimum, obie pochodne cząstkowe tej funkcji przyrównujemy do zera:

$$\begin{cases} \frac{\partial f}{\partial b_0} = \sum_{i=1}^n 2(y_i - (b_0 + b_1 x_i))(-1) = 0 \\ \frac{\partial f}{\partial b_1} = \sum_{i=1}^n 2(y_i - (b_0 + b_1 x_i))(-x_i) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n (y_i - (b_0 + b_1 x_i)) = 0 \\ \sum_{i=1}^n (y_i - (b_0 + b_1 x_i))x_i = 0. \end{cases} \quad (1)$$

Układ równań oznaczony (1) nazywamy równaniami normalnymi. Dzieląc oba te równania obustronnie przez n otrzymujemy

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n y_i - b_0 - b_1 \frac{1}{n} \sum_{i=1}^n x_i = 0 \\ \frac{1}{n} \sum_{i=1}^n x_i y_i - b_0 \frac{1}{n} \sum_{i=1}^n x_i - b_1 \frac{1}{n} \sum_{i=1}^n x_i^2 = 0. \end{cases}$$

Oznaczając $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} := \frac{1}{n} \sum_{i=1}^n y_i$, $\overline{xy} := \frac{1}{n} \sum_{i=1}^n x_i y_i$, $\bar{x}^2 := \frac{1}{n} \sum_{i=1}^n x_i^2$, powyższy układ równań możemy przepisać w następującej postaci

$$\begin{cases} \bar{y} - b_0 - b_1 \bar{x} = 0 \\ \overline{xy} - b_0 \bar{x} - b_1 \bar{x}^2 = 0 \end{cases} \Leftrightarrow \begin{cases} b_0 = \bar{y} - b_1 \bar{x} \\ \overline{xy} - (\bar{y} - b_1 \bar{x})\bar{x} - b_1 \bar{x}^2 = 0 \end{cases} \Leftrightarrow \begin{cases} b_0 = \bar{y} - b_1 \bar{x} \\ \overline{xy} - \bar{y}\bar{x} = b_1(\bar{x}^2 - (\bar{x})^2). \end{cases}$$

Zatem otrzymujemy, że rozwiązaniem układu równań (1) jest

$$\begin{cases} \hat{b}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \\ \hat{b}_0 = \bar{y} - \hat{b}_1\bar{x} \end{cases}, \quad (2)$$

Można pokazać, że funkcja f w punkcie (\hat{b}_0, \hat{b}_1) danym (2) rzeczywiście osiąga wartość najmniejszą. Stąd szukane wartości estymatorów β_0 i β_1 to

$$\begin{cases} \hat{\beta}_1 = \frac{\overline{xY} - \bar{x}\bar{Y}}{\overline{x^2} - (\bar{x})^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}. \end{cases}$$

Sposób II - metoda największej wiarygodności

Polega ona na dobraniu $\hat{\beta}_0, \hat{\beta}_1$ i $\hat{\sigma}_{MNW}^2$ tak by funkcja wiarygodności (tzn. łączna gęstość wektora losowego (Y_1, Y_2, \dots, Y_n)) dla tych dobranych wartości osiągała wartość największą.

Po przeprowadzeniu odpowiednich rachunków otrzymujemy

$$\begin{cases} \hat{\beta}_1 = \frac{\overline{xY} - \bar{x}\bar{Y}}{\overline{x^2} - (\bar{x})^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x} \\ \hat{\sigma}_{MNW}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2. \end{cases}$$

Zauważmy, że otrzymaliśmy te same estymatory parametrów β_0 i β_1 co metodą najmniejszych kwadratów. Ponadto można pokazać, że estymatory te są nieobciążone, tzn. $E(\hat{\beta}_0) = \beta_0$ i $E(\hat{\beta}_1) = \beta_1$. Natomiast estymator $\hat{\sigma}_{MNW}^2$ jest estymatorem obciążonym, dlatego modyfikujemy go tak by stał się nieobciążony - estymatorem nieobciążonym jest

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2$$

i właśnie tego nieobciążonego estymatora $\hat{\sigma}^2$ używa się do szacowania nieznanej wariancji błędów w modelu regresji liniowej.

Wprowadźmy następujące nazwy i oznaczenia.

- Prostą $y = \hat{b}_0 + \hat{b}_1x$, gdzie \hat{b}_0 i \hat{b}_1 zostały wyznaczone metodą najmniejszych kwadratów (tzn. dane są wzorem (2)), nazywamy *prostą regresji*.
- Wielkość $\hat{y}_i := \hat{b}_0 + \hat{b}_1x_i$ nazywamy *wartością prognozowaną* lub *prognozą* dla i -tej obserwacji.
- Wielkość $e_i := y_i - \hat{y}_i = y_i - (\hat{b}_0 + \hat{b}_1x_i)$ nazywamy i -tym *błędem* lub *wartością resztową* lub *residuum*.

Zauważmy, że z (1) otrzymujemy

$$\sum_{i=1}^n e_i = 0 \quad \text{ i } \quad \sum_{i=1}^n e_i x_i = 0$$

a stąd

$$\sum_{i=1}^n e_i \hat{y}_i = \sum_{i=1}^n e_i (\hat{b}_0 + \hat{b}_1 x_i) = \hat{b}_0 \sum_{i=1}^n e_i + \hat{b}_1 \sum_{i=1}^n e_i x_i = \hat{b}_0 \cdot 0 + \hat{b}_1 \cdot 0 = 0. \quad (3)$$

Współczynnik determinacji R^2

Tak zwaną *całkowitą sumę kwadratów SST*, czyli sumę kwadratów odpowiadającą za całkowitą zmienność zmiennej objaśnianej y :

$$SST := \sum_{i=1}^n (y_i - \bar{y})^2$$

możemy rozbić na dwa składniki:

$$\begin{aligned}\sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2,\end{aligned}$$

ponieważ na mocy (3) mamy

$$\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^n e_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^n e_i \hat{y}_i - \bar{y} \sum_{i=1}^n e_i = 0 - \bar{y} \cdot 0 = 0.$$

Oznaczając

$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ - suma kwadratów błędów (odpowiada za zmienność y nie wyjaśnioną przez prostą regresji),

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ - regresyjna suma kwadratów (odpowiada za zmienność y objaśnianą przez prostą regresji),

otrzymujemy

$$SST = SSE + SSR.$$

Współczynnik determinacji R^2 definiujemy następująco:

$$R^2 := \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

a jego wartość interpretujemy jako stosunek zmienności y objaśnianej przez prostą regresji (czyli zmienności y objaśnianej przez x) do zmienności całkowitej y . Innymi słowy, $R^2 \cdot 100\%$ to procent zmienności y , która jest objaśniana przez model (objaśniana przez x).

Można pokazać, że $R^2 = r^2$, gdzie r to współczynnik korelacji próbkowej między x_i i y_i , tzn. $r = \frac{1}{n-1} \sum_{i=1}^n \frac{x_i - \bar{x}}{S_x} \frac{y_i - \bar{y}}{S_y}$, gdzie $S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ i $S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$.

2 Regresja liniowa wielokrotna

Teraz rozważamy sytuację, w której mamy zmienną objaśnianą Y i p zmiennych objaśniających x_1, x_2, \dots, x_p . Dane będziemy mieć zatem zapisane w postaci n wektorów $(p+1)$ -wymiarowych:

$$(x_{11}, x_{12}, \dots, x_{1p}, y_1), (x_{21}, x_{22}, \dots, x_{2p}, y_2), \dots, (x_{n1}, x_{n2}, \dots, x_{np}, y_n),$$

gdzie x_{ij} oznacza wartość j -tej zmiennej objaśniającej dla i -tego obiektu.

Np. możemy badać zależność rekordowego czasu pokonania trasy w biegu przełajowym od długości tej trasy i od różnicy poziomów do pokonania na tej trasie. Wtedy będziemy potrzebować dane następującej postaci:

$$(x_{11}, x_{12}, y_1), (x_{21}, x_{22}, y_2), \dots, (x_{n1}, x_{n2}, y_n),$$

gdzie

- y_i oznacza rekordowy czas pokonania i -tej trasy;
- x_{i1} oznacza długość i -tej trasy;
- x_{i2} oznacza różnicę poziomów do pokonania na i -tej trasie.

Do tych danych będziemy próbować dopasować model liniowy

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

gdzie ε_i to niezależne zmienne losowe o rozkładzie $\mathcal{N}(0, \sigma^2)$.

Ogólnie **model wielokrotnej regresji liniowej** jest następujący

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n, \\ \text{gdzie } \varepsilon_i &\text{ to niezależne zmienne losowe o rozkładzie } \mathcal{N}(0, \sigma^2), \\ \text{natomiast } \beta_0, \beta_1, \dots, \beta_p &\text{ i } \sigma^2 > 0 \text{ to pewne stałe rzeczywiste (tzw. parametry modelu).} \end{aligned}$$

Oznaczając

$$\mathbb{Y} = \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{wektor odpowiedzi}}, \quad \mathbb{X} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}}_{\text{macierz eksperymentu}}, \quad \mathbb{B} = \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\text{wektor parametrów}}, \quad \mathbb{E} = \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\text{wektor błędów}},$$

powyższy model możemy zapisać w postaci macierzowej

$$\mathbb{Y} = \mathbb{X} \cdot \mathbb{B} + \mathbb{E}.$$

Metodą najmniejszych kwadratów otrzymujemy następujący estymator wektora parametrów

$$\hat{\mathbb{B}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (\mathbb{X}^T \cdot \mathbb{X})^{-1} \cdot \mathbb{X}^T \cdot \mathbb{Y},$$

przy założeniu, że macierz $\mathbb{X}^T \cdot \mathbb{X}$ jest odwracalna, co zachodzi, gdy kolumny macierzy \mathbb{X} są liniowo niezależne. Stosując metodę największej wiarygodności otrzymujemy dokładnie ten sam estymator $\hat{\mathbb{B}}$. Ponadto można pokazać, że $\hat{\mathbb{B}}$ jest estymatorem nieobciążonym wektora parametrów \mathbb{B} , tzn. $E(\hat{\mathbb{B}}) = \mathbb{B}$.

Do szacowania nieznanej wariancji błędów σ^2 używamy nieobciążonego estymatora tej wielkości, danego wzorem

$$\hat{\sigma}^2 = \frac{1}{n - (p + 1)} \sum_{i=1}^n e_i^2,$$

gdzie $e_i = Y_i - \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip})}_{\text{ozn. } \hat{Y}_i - \text{prognoza dla } i\text{-tej obserwacji}}$ to i -ty błąd (*residuum, wartość resztowa*).

Stąd

$$\hat{\sigma}^2 = \frac{1}{n - (p + 1)} \sum_{i=1}^n \left(Y_i - \sum_{j=0}^p x_{ij} \hat{\beta}_j \right)^2 = \frac{1}{n - (p + 1)} (\mathbb{Y} - \mathbb{X} \cdot \hat{\mathbb{B}})^T \cdot (\mathbb{Y} - \mathbb{X} \cdot \hat{\mathbb{B}}),$$

gdzie $\begin{bmatrix} x_{10} & x_{20} & \dots & x_{n0} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$.

Zauważmy, że wektor prognoz $\hat{\mathbb{Y}} = \begin{bmatrix} \hat{Y}_1 & \hat{Y}_2 & \dots & \hat{Y}_n \end{bmatrix}^T$ dany jest wzorem

$$\hat{\mathbb{Y}} = \mathbb{X} \cdot \hat{\mathbb{B}} = \underbrace{\mathbb{X} \cdot (\mathbb{X}^T \cdot \mathbb{X})^{-1} \cdot \mathbb{X}^T}_{\text{ozn. } H} \cdot \mathbb{Y} = H \cdot \mathbb{Y}.$$

Macierz H nazywana jest *macierzą daszkową* (bo przekształca \mathbb{Y} na $\hat{\mathbb{Y}}$ czyli na \mathbb{Y} z daszkiem).

3 Testy w modelu regresji liniowej

1. Testy istotności dla poszczególnych współczynników

Ustalamy $i_0 \in \{0, 1, \dots, p\}$. Weryfikujemy hipotezę

$$H_0 : \beta_{i_0} = 0 \quad (i_0\text{-ta zmienna nie jest istotna w sytuacji, gdy wszystkie inne zmienne objaśniające pozostają w modelu; gdy } i_0 = 0 \text{ oznacza to, że wyraz wolny nie jest w modelu istotny})$$

przeciwko hipotezie

$$H_1 : \beta_{i_0} \neq 0.$$

Statystyka testowa

$$T = \frac{\hat{\beta}_{i_0}}{SE(\hat{\beta}_{i_0})}, \text{ gdzie } SE(\hat{\beta}_{i_0}) \text{ jest estymatorem odchylenia standardowego } \hat{\beta}_{i_0},$$

w sytuacji, gdy H_0 jest prawdziwa, ma rozkład t -Studenta o $n - (p + 1)$ stopniach swobody. Estymator $SE(\hat{\beta}_{i_0})$ wyznaczamy jako pierwiastek z i_0 -tego elementu na przekątnej macierzy $\hat{\sigma}^2(\mathbb{X}^T \cdot \mathbb{X})^{-1}$.

Zbiór krytyczny to

$$W = (-\infty, -t_{1-\frac{\alpha}{2}, n-(p+1)}] \cup [t_{1-\frac{\alpha}{2}, n-(p+1)}, +\infty),$$

gdzie $t_{\beta, m}$ oznacza kwantyl rzędu β rozkładu t -Studenta o m stopniach swobody.

Jeśli $T \in W$, to na poziomie istotności α , H_0 odrzucamy.

2. Test F czy którakolwiek ze zmiennych objaśniających jest istotna

$$\begin{aligned} H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 & \quad (\text{żadna zmienna w modelu nie jest istotna}) \\ H_1 : \text{istnieje } i \neq 0 \text{ takie, że } \beta_i \neq 0 & \quad (\text{co najmniej jedna zmienna w modelu jest istotna}) \end{aligned}$$

Statystyka testowa

$$F = \frac{SSR/p}{SSE/(n - (p + 1))}$$

w sytuacji, gdy H_0 jest prawdziwa, ma rozkład F -Snedecora o p i $n - (p + 1)$ stopniach swobody.

Zbiór krytyczny to

$$W = [f_{1-\alpha, p, n-(p+1)}, +\infty),$$

gdzie $f_{\beta, m, r}$ oznacza kwantyl rzędu β rozkładu F -Snedecora o m i r stopniach swobody.

Jeśli $F \in W$, to na poziomie istotności α , H_0 odrzucamy.

Implementacja testów z 1. i 2. w R:

```
> model.liniowy=lm(zm.objasniana~ zm.objasniajaca.0+...+zm.objasniajaca.p)
> summary(model.liniowy)
```

3. Test czy pewien podzbiór zmiennych objaśniających jest istotny (tzw. częściowy test F)

Rozważamy dwa modele liniowe: model mniejszy zawarty w modelu większym:

$$\text{model mniejszy (m.m.): } Y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$$\text{model większy (m.w.): } Y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \beta_{p+1} x_{i, p+1} + \dots + \beta_q x_{iq} + \tilde{\varepsilon}_i,$$

gdzie $p < q$.

Weryfikujemy hipotezę

$$H_0 : \beta_{p+1} = \dots = \beta_q = 0 \quad (\text{model mniejszy jest poprawny})$$

przeciwko hipotezie

$$H_1 : \text{istnieje } i \in \{p + 1, \dots, q\} \text{ takie, że } \beta_i \neq 0 \quad (\text{potrzebny jest model większy})$$

Statystyka testowa

$$F = \frac{(SSE_{m.m.} - SSE_{m.w.})/(q-p)}{SSE_{m.w.}/(n-(q+1))}$$

w sytuacji, gdy H_0 jest prawdziwa, ma rozkład F -Snedecora o $q-p$ i $n-(q+1)$ stopniach swobody.

Zbiór krytyczny to

$$W = [f_{1-\alpha, q-p, n-(q+1)}, +\infty),$$

gdzie, jak poprzednio, $f_{\beta, m, r}$ oznacza kwantyl rzędu β rozkładu F -Snedecora o m i r stopniach swobody.

Jeśli $F \in W$, to na poziomie istotności α , H_0 odrzucamy.

Implementacja w R:

```
> model.mniejszy=lm(zm.objasniana~ zm.objasnajaca.0+...+zm.objasnajaca.p)
> model.wiekszy=lm(zm.objasniana~ zm.objasnajaca.0+...+zm.objasnajaca.q)
> anova(model.mniejszy, model.wiekszy)
```

4 Diagnostyka dopasowania modelu regresji liniowej

Analizowanie jedynie wartości współczynnika determinacji R^2 i wyników testów w modelu regresji, nie wystarcza by prawidłowo ocenić czy dobry model został wybrany do opisu danych. Poprawność testów bowiem, jak i poprawność prognoz wyznaczanych na podstawie modelu, zależą w istotny sposób od poprawności postulowanego modelu. Przypominajmy, że w modelu regresji liniowej zakładamy, że

- dla ustalonego \mathbb{X} , $E(\mathbb{Y}) = \mathbb{B} \cdot \mathbb{X}$;
- błędy ε_i , $i = 1, 2, \dots, n$, są niezależne o tym samym rozkładzie $\mathcal{N}(0, \sigma^2)$ (wszczegółności zakładamy, że błędy mają równe wariancje).

Aby sprawdzić założenia dotyczące błędów ε_i , $i = 1, 2, \dots, n$, analizujemy residua $e_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$. Residua przybliżają błędy; im większe n , tym przybliżenie jest lepsze. Zatem, jeśli model jest poprawny, to residua powinny w przybliżeniu zachowywać się jak niezależne zmienne losowe o tym samym rozkładzie normalnym $\mathcal{N}(0, \sigma^2)$ (w szczególności jak zmienne losowe o równych wariancjach).

1. Rysujemy

- wykres residuów w funkcji x : (x_i, e_i) jeśli mamy tylko jedną zmienną objaśniającą lub wykresy residuów w funkcji kolejnych zmiennych objaśniających, gdy tych zmiennych jest więcej niż jedna;
- lub
- wykres residuów w funkcji numeru porządkowego: (i, e_i) ;
- lub
- wykres residuów w funkcji prognoz: (\hat{Y}_i, e_i) .

Wszystkie te wykresy powinny przedstawiać chmurę punktów skupioną wokół osi OX , nie mającą wyraźnej struktury ani tendencji.

2. Jeśli n nie jest duże, to lepszym rozwiązaniem, niż analizowanie residuów, jest patrzenie na tzw. *residua standardyzowane*

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}},$$

gdzie h_{ii} to i -ty element z przekątnej macierzy daszkowej H . Wynika to stąd, że residua nie muszą mieć równych wariancji, nawet w sytuacji, gdy wariancje błędów są równe:

$$\text{Var}(e_i) = \sigma^2(1-h_{ii}).$$

Po podzieleniu e_i przez $\hat{\sigma}\sqrt{1-h_{ii}}$ (czyli przez estymator odchylenia standardowego e_i) otrzymujemy $Var(r_i) \approx 1$, $i = 1, 2, \dots, n$.

Poniższa funkcja w R:

```
> plot(nazwa.modelu liniowego)
```

wykonuje 4 wykresy:

- (a) wykres residuów w funkcji prognoz,
 - (b) wykres kwantylowy dla residuów standardyzowanych,
 - (c) wykres $\sqrt{|r_i|}$ w funkcji prognoz (niektórzy właśnie taki wykres zalecają by wykryć ewentualne nierówne wariancje błędów, pierwiastek ma zredukować asymetrię pojawiającą się z powodu wartości bezwzględnej),
 - (d) o 4-tym wykresie będzie mowa później.
3. Szukamy *obserwacji wpływowych* czyli obserwacji, które znacznie wpływają na dopasowany model (obrazowo mówiąc - przyciągają dopasowaną hiperpłaszczyznę).

Jak wykryć obserwacje wpływowe?

Dla obserwacji wpływowej \hat{Y}_i znacznie zależy od Y_i . Zauważmy, że

$$\hat{\mathbb{Y}} = H \cdot \mathbb{Y} \Rightarrow \hat{Y}_i = h_{i1}Y_1 + h_{i2}Y_2 + \dots + h_{ii}Y_i + \dots + h_{in}Y_n = \sum_{j=1, j \neq i}^n h_{ij}Y_j + h_{ii}Y_i.$$

Widać, że im h_{ii} większe, tym \hat{Y}_i bardziej zależy od Y_i . Zatem za miarę wpływu Y_i na \hat{Y}_i można przyjąć h_{ii} - i -ty element z przekątnej macierzy daszkowej H . Ponadto dowodzi się, że

$$\frac{1}{n} \leq h_{ii} \leq 1 \text{ dla każdego } i = 1, 2, \dots, n, \text{ oraz}$$

$$\sum_{i=1}^n h_{ii} = p + 1.$$

Uznaje się, że jeśli

$$h_{ii} > 2 \cdot \text{średnie}(h_{ii}) = 2 \cdot \frac{p+1}{n},$$

to i -ta obserwacja jest potencjalną obserwacją wpływową.

4. Szukamy *obserwacji odstających* czyli takich, które nie pasują do wzorca sugerowanego przez pozostałe punkty.

W przypadku obserwacji odstającej

$$r_i = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$$

będzie duże.

Uznaje się, że obserwacja jest odstająca gdy $|r_i| > 2$, zmieniając ten warunek na $|r_i| > 4$, gdy mamy bardzo duży zbiór danych.

Powyższa reguła może jednak nie wychwycić wpływowych obserwacji odstających, bo dla nich $Y_i - \hat{Y}_i$ może być bardzo małe. Aby zidentyfikować także takie obserwacje odstające, analizuje się *residua modyfikowane*

$$d_i = Y_i - \hat{Y}_{i(i)},$$

gdzie $\hat{Y}_{i(i)}$ to prognoza dla Y_i na podstawie modelu regresji liniowej wyznaczonego z pominięciem i -tej obserwacji.

5. Do wykrycia obserwacji wpływowych i odstających służy także miara zwana odległością Cook'a. *Odległość Cooke'a* D_i to miara wpływu jaki ma i -ta obserwacja na dopasowaną hiperpłaszczyznę:

$$D_i = \frac{\sum_{j=1}^n (\hat{Y}_{j(i)} - \hat{Y}_j)^2}{(p+1)\hat{\sigma}^2} = \frac{(\hat{\mathbb{B}} - \hat{\mathbb{B}}_{(i)})^T \cdot \mathbb{X}^T \cdot \mathbb{X} \cdot (\hat{\mathbb{B}} - \hat{\mathbb{B}}_{(i)})}{(p+1)\hat{\sigma}^2},$$

gdzie

- $\hat{Y}_{j(i)}$ to wartość przewidywana dla j -tej obserwacji, obliczona na podstawie modelu regresji liniowej wyznaczonego z pominięciem i -tej obserwacji;
- $\hat{\mathbb{B}}_{(i)}$ to estymator parametrów modelu regresji liniowej na podstawie danych, z których usunięto i -tą obserwację.

Duża wartość D_i wskazuje na to, że usunięcie i -tej obserwacji ma znaczny wpływ na prognozy znanych wartości zmiennej objaśnianej.

Można pokazać, że

$$D_i = \frac{1}{p+1} \cdot r_i^2 \cdot \frac{h_{ii}}{1-h_{ii}},$$

skąd widać, że wartość D_i będzie duża, gdy $|r_i|$ będzie duże (a duża wartość $|r_i|$ sugeruje, że i -ta obserwacja jest obserwacją odstającą) lub gdy h_{ii} będzie duże (czyli bliskie 1, co sugeruje, że i -ta obserwacja jest obserwacją wpływową). W literaturze zaleca się by uznawać, że D_i nie jest duże, gdy jest znacząco mniejsze od 1, ale wskazane jest patrzeć nie tylko na samą wartość D_i , ale także na to jakie są odstępstwa pomiędzy uporządkowanymi wartościami D_i - duże skoki powinny wzbudzać nasze zainteresowanie.

Czwarty wykres generowany przez komendę

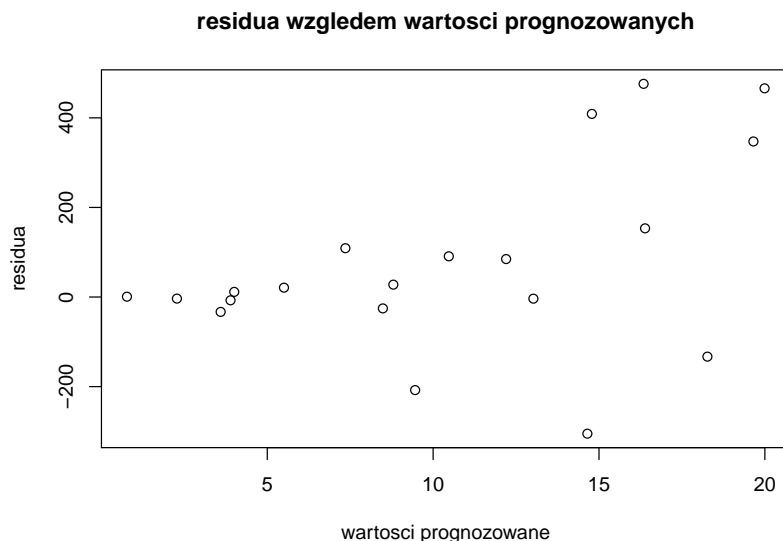
```
> plot(nazwa.modelu.liniowego)
```

to wykres standaryzowanych residuów r_i w zależności od wpływów h_{ii} . Na wykresie tym naniesione są czerwone, przerywane krzywe (jeśli tylko się mieszczą) odpowiadające wartościom $D_i = 0.5$ i $D_i = 1$. Obserwacje wymagające bliższego przyjrzenia się to obserwacje o dużej odległości Cook'a D_i .

Jak postępować w przypadku wykrycia obserwacji wpływowych lub odstających?

- Jeśli obserwacja ta jest nietypowa i w pewien sposób różni się od pozostałych danych, warto spróbować ją usunąć i na nowo dopasować model regresji.
- Nie należy jednak usuwać takich obserwacji automatycznie i bezmyślnie. Jeśli taka obserwacja nie pojawiła się w wyniku błędu i nie widzimy by była ona nietypowa, to warto próbować dopasować inny model; np. dodać zmienne objaśniające (dołączyć np. kolejne potęgi zmiennej objaśniającej) lub próbować przekształcać zmienną objaśnianą lub zmienne objaśniające.

6. Jednym z założeń modelu regresji liniowej jest to, że błędy mają równe wariancje. Aby sprawdzić czy założenie to jest spełnione, analizujemy wykresy residuów, o których była mowa w pkt 1.



Na powyższym rysunku wariancja residuów rośnie wraz ze wzrostem wartości prognozowanych. Świadczy to o złym dopasowaniu modelu do danych.

Jak postępować w sytuacji, gdy uznamy, że założenie o równych wariancjach błędów nie jest spełnione? Mamy dwie możliwości rozwiązania tego problemu.

- Zastosować *metodę najmniejszych ważonych kwadratów* do wyznaczenia współczynników modelu regresji liniowej, tzn. za estymator wektora parametrów \mathbb{B} przyjąć wektor $\hat{\mathbb{B}}_W = [\hat{\beta}_{0W}, \hat{\beta}_{1W}, \dots, \hat{\beta}_{pW}]$, który minimalizuje sumę

$$\sum_{i=1}^n w_i (y_i - (\beta_{0W} + \beta_{1W}x_{i1} + \dots + \beta_{pW}x_{ip}))^2,$$

gdzie waga w_i powinna być tym mniejsza im $Var(\varepsilon_i)$ jest większa. W praktyce przyjmuje się $w_i = \frac{1}{\hat{\sigma}_i^2}$, gdzie $\hat{\sigma}_i^2$ jest pewnym estymatorem lub oszacowaniem $Var(\varepsilon_i) = Var(Y_i)$. Zatem metodę tą możemy w praktyce zastosować, gdy jesteśmy w stanie oszacować lub wyestymować $Var(Y_i)$. Możemy to zrobić np. w przypadku, w którym Y_i to średnie policzone na podstawie n_i obserwacji. Wtedy $Var(Y_i)$ są proporcjonalne do $\frac{1}{n_i}$, zatem możemy przyjąć $w_i = n_i$.

Implementacja w R:

```
> lm(zm.objasniana ~ zm.objasnajaca.1 + ... + zm.objasnajaca.p, weights =  $\underbrace{\hspace{1cm}}$  )
```

wektor z wagami

- Przekształcić zmienną objaśnianą Y lub zmienne objaśniane tak by po przekształceniu wariancje błędów stały się w przybliżeniu równe.
 - Jeśli któraś ze zmiennych jest typu zliczającego, to spierwiastkowanie jej często okazuje się dobrym wyborem.
 - Inne, często rozwiązujące problem przekształcenie, to logarytmowanie.
 - Jeśli zmienna objaśniana i zmienna objaśniająca (zmienne objaśniające) mają tę samą jednostkę, to zalecane jest poddawanie ich temu samemu przekształceniu.

7. Przekształcenia zmiennych stosuje się także wtedy, gdy błędy nie mają rozkładu normalnego.

- Do znajdowania przekształcenia dodatniej zmiennej odpowiedzi Y , dającego najlepiej dopasowany model, stosuje się *metodę Boxa-Coxa*. Przekształcenie jest wybierane z rodziny przekształceń

$$g_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{gdy } \lambda \neq 0 \\ \ln y & \text{gdy } \lambda = 0. \end{cases}$$

Parametr λ dobieramy tak, by maksymalizował funkcję wiarygodności. Metoda Boxa-Coxa jest wrażliwa na obserwacje odstające - gdy otrzymamy $\hat{\lambda} = 5$, to właśnie one najprawdopodobniej będą powodem tak dużej wartości λ .

Implementacja w R:

```
> library(MASS)
> boxcox(nazwa.modelu.liniowego, lambda =  $\underbrace{\hspace{1cm}}$  )
```

wektor z wartościami λ ,
które chcemy rozpatrzyć

Domyślnie mamy $\text{lambda} = \text{seq}(-2, 2, \text{by} = 0.1)$, więc jeśli chcemy rozpatrzyć $\lambda = -2; -1, 9; -1, 8; \dots; 1, 9; 2$, to wystarczy napisać

```
> boxcox(nazwa.modelu.liniowego)
```

- Gdy w modelu mamy tylko jedną zmienną objaśniającą, w celu oceny jej wpływu na zmienną objaśnianą, patrzymy na wykres y_i w funkcji x_i . Wykres ten może zasugerować stosowną transformację zmiennej objaśniającej lub dodanie do modelu kolejnych potęg zmiennej objaśniającej.
- Gdy mamy więcej zmiennych objaśniających, w celu wizualizacji wpływu jednej wybranej zmiennej objaśniającej x_{i_0} na zmienną objaśnianą, po usunięciu liniowego wpływu pozostałych zmiennych objaśnianych, można użyć *częściowych wykresów regresji* (*partial regression plots* lub *partial leverage plots*).

8. Współliniowość zmiennych objaśniających.

Jeśli macierz $\mathbb{X}^T \cdot \mathbb{X}$ jest nieodwracalna (czyli jeśli zmienne objaśniające są liniowo zależne), to estymator $\hat{\mathbb{B}}$ nie jest wyznaczony jednoznacznie.

Np. jeśli w modelu regresji

$$\mathbb{Y} = 1 + 3x_1 + 5x_2 + \varepsilon \quad \text{mamy} \quad x_2 = 2x_1,$$

to możemy go równoważnie przepisać np. jako $\mathbb{Y} = 1 + 7x_1 + 3x_2 + \varepsilon$ i estymator $\hat{\mathbb{B}} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]^T$ może być zarówno bliski $[1, 3, 5]^T$ jak i $[1, 7, 3]^T$. Gdy macierz $\mathbb{X}^T \cdot \mathbb{X}$ jest bliska macierzy nieodwracalnej (czyli występuje przybliżona zależność liniowa między zmiennymi objaśniającymi), to choć $\hat{\mathbb{B}}$ jest wyznaczona jednoznacznie, jej zmienność może być bardzo duża.

Można pokazać, że w modelu liniowym z p zmiennymi objaśniającymi

$$\text{Var}(\hat{\beta}_j) = \frac{1}{1 - R_j^2} \cdot \frac{\sigma^2}{(n - 1)S_{x_j}^2}, \quad j = 1, 2, \dots, p,$$

gdzie R_j^2 to współczynnik determinacji dla modelu liniowego, w którym x_j to zmienna objaśniana a pozostałe x_i , $i \neq j$, to zmienne objaśniające (gdy w modelu mamy tylko dwie zmienne, jest to kwadrat współczynnika korelacji między tymi zmiennymi, zatem R_j możemy traktować jako uogólnienie współczynnika korelacji na więcej zmiennych).

Czynnik $\frac{1}{1 - R_j^2}$ jest zwany *współczynnikiem podbicia wariancji* VIF_j (*variance inflation factor*):

$$\text{VIF}_j = \frac{1}{1 - R_j^2}.$$

Duża wartość VIF_j (w praktyce $\text{VIF}_j > 5$ jest uznawane za duże a $\text{VIF}_j > 10$ za bardzo duże) sugeruje, że zmienna x_j jest silnie zależna liniowo od pozostałych zmiennych i warto próbować usunąć ją z modelu.

5 Selekcja zmiennych w modelu regresji liniowej

Przed przystąpieniem do selekcji zmiennych należy:

- Zastosować wszelkie transformacje danych, które wydają się być stosowne.
- Zidentyfikować obserwacje odstające i wpływowe (bo metody selekcji zmiennych są na nie nieodporne) i być może wyłączyć je z analizy, przynajmniej tymczasowo. Jeśli nie będziemy mieli podstaw do usunięcia tych obserwacji, należy zastosować metody selekcji zmiennych oparte na regresji odpornej (tych metod nie będziemy omawiać).

Kryteria służące do wyboru najlepszego modelu:

1. Współczynnik determinacji R^2 i skorygowany współczynnik determinacji R_{adj}^2

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},$$

gdzie, jak w modelu regresji jednokrotnej,

$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$ - suma kwadratów odpowiadająca za całkowitą zmienność zmiennej objaśnianej Y

$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$ - suma kwadratów błędów (odpowiada za zmienność Y nie wyjaśnioną przez model regresji),

$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ - regresyjna suma kwadratów (odpowiada za zmienność Y objaśnianą przez model regresji)

i zachodzi $SST = SSE + SSR$.

Tak jak w przypadku regresji jednokrotnej, wartość współczynnika determinacji R^2 w modelu wielokrotnej regresji liniowej interpretujemy jako stosunek zmienności Y objaśnianej przez model do zmienności całkowitej Y .

Jeśli do modelu dodajemy jedną lub więcej zmiennych objaśniających, to R^2 nie maleje (zazwyczaj wzrasta), nawet jeśli dodana zmienna/zmienne są w modelu nieistotne. Zatem, porównując dopasowanie modeli liniowych o różnej liczbie zmiennych objaśniających, należy patrzeć na wartość skorygowanego współczynnika determinacji R_{adj}^2 , a nie na R^2 :

$$R_{adj}^2 = 1 - \frac{SSE/(n - (p + 1))}{SST/(n - 1)}.$$

Można wykazać, że dodanie zmiennej objaśniającej (lub zmiennych objaśniających) do modelu regresji zwiększy R_{adj}^2 jedynie, gdy wartość statystyki częściowego testu F sprawdzającego istotność tych zmiennych będzie większa od 1.

2. Kryterium Akaike AIC (*Akaike Information Criterion*)

$$\begin{aligned} \text{AIC} &= -2[\text{funkcja wiarygodności} - (p + 2)] \\ &= n \ln\left(\frac{SSE}{n}\right) + 2p + \underbrace{(\text{stała})}_{\text{można pominąć}}, \end{aligned}$$

gdzie (stała) nie zależy od SSE i p , czyli jest taka sama dla wszystkich modeli zbudowanych na podstawie tego samego zbioru danych, więc może zostać pominięta podczas porównań.

Im AIC przyjmuje mniejszą wartość, tym model jest lepszy.

Gdy n jest małe lub gdy $\frac{n}{p+2} \leq 40$, to AIC ma tendencję do preferowania modeli z za dużą liczbą zmiennych. Dlatego w tej sytuacji zalecane jest używanie skorygowanego kryterium Akaike:

$$\text{AIC}_c = \text{AIC} + \frac{2(p + 2)(p + 3)}{n - p - 1}.$$

3. Kryterium Schwarza BIC (*Schwarz Criterion* lub *Bayes Information Criterion*)

$$\begin{aligned} \text{BIC} &= n \ln\left(\frac{SSE}{n}\right) + (p + 2) \ln n \\ &= n \ln\left(\frac{SSE}{n}\right) + p \ln n + \underbrace{2 \ln n}_{\text{można pominąć}}, \end{aligned}$$

gdzie $2 \ln n$ jest stałą niezależną od SSE i p , czyli jest takie samo dla wszystkich modeli zbudowanych na podstawie tego samego zbioru danych, więc może zostać pominięte podczas porównań.

Podobnie jak w przypadku AIC, im BIC przyjmuje mniejszą wartość, tym model jest lepszy.

Chcemy wybrać „najlepszy” podzbiór zmiennych objaśniających, usuwając zmienne nieistotne. Poniższe trzy metody wyboru takiego podzbioru możemy przeprowadzić opierając się na jednym z kryteriów: R_{adj}^2 , AIC, AIC_c lub BIC.

1. Metoda eliminacji (*backward elimination*)

Rozpoczynamy od modelu pełnego, zawierającego wszystkie zmienne objaśniające. W każdym kroku usuwamy zmienną, która ma największą wartość p -value testu istotności tej jednej zmiennej (ponieważ w ustalonym kroku porównujemy modele o tej samej liczbie zmiennych, wybór zmiennej o największym p -value jest równoważny wyborowi zmiennej, po usunięciu której otrzymamy model o najmniejszej wartości kryterium informacyjnego lub równoważnie o największej wartości R_{adj}). Proces ten kontynuujemy aż do momentu, gdy kryterium wskazuje, że po usunięciu zmiennej o największym p -value model ulegnie pogorszeniu (R_{adj}^2 zmaleje albo AIC, AIC_c albo BIC wzrośnie) lub, gdy usuniemy wszystkie zmienne z modelu.

Warunek zakończenia procedury można także oprzeć na samej wartości p -value - kończyć procedurę, gdy usuniemy wszystkie zmienne z modelu lub gdy wartość p -value testu istotności dla zmiennej wybranej do usunięcia, jest mniejsza bądź równa pewnej wartości granicznej p_{OUT} (zwykle przyjmuje się $p_{OUT} = 0.05$ lub $p_{OUT} = 0.1$).

2. Metoda dołączania (*forward selection*)

Rozpoczynamy od modelu zawierającego jedynie wyraz wolny. W każdym kroku dodajemy zmienną, która ma najmniejszą wartość p -value testu istotności tej jednej zmiennej. Proces ten kontynuujemy aż do momentu, gdy kryterium wskazuje, że po dodaniu zmiennej o najmniejszym p -value model ulegnie pogorszeniu (R_{adj}^2 zmaleje albo AIC, AIC_c albo BIC wzrośnie) lub, gdy dodamy wszystkie zmienne do modelu.

Podobnie jak w przypadku metody eliminacji, warunek zakończenia procedury metody dołączania można też oprzeć na samej wartości p -value - kończyć procedurę, gdy dodamy wszystkie zmienne do modelu lub gdy wartość p -value testu istotności dla zmiennej wybranej do dodania, jest większa bądź równa pewnej wartości granicznej p_{IN} (zwykle przyjmuje się $p_{IN} = 0.1$ lub $p_{IN} = 0.2$).

3. Metoda selekcji krokowej (*stepwise regression*)


Rozpoczynamy od modelu zawierającego jedynie wyraz wolny. Pierwszy krok jest krokiem z metody dołączania, po nim wykonujemy krok z metody eliminacji i tak na zmianę, aż do momentu, gdy stwierdzimy, że nie dodajemy żadnej zmiennej do modelu.

Implementacja w R:

```
> step(object=model.pocztakowy, scope, direction=c("backward", "forward", "both"), k)
```

Jako `object` podajemy model, z którym rozpoczynamy procedurę (czyli model pełny w przypadku metody eliminacji oraz model zawierający jedynie wyraz wolny w przypadku metody dołączania i metody selekcji krokowej). Natomiast jako `scope` podajemy listę zawierającą modele, które chcemy brać pod uwagę. Jeśli przyjmiemy

- $k=2$ (tak jest domyślnie), to wybór najlepszego modelu będzie oparty na kryterium AIC;
- $k=\log(\text{length}(\text{zmienna.objaśniana}))$, to wybór najlepszego modelu będzie oparty na kryterium BIC.

```
> model.pelny=lm(zm.objasniana ~ ., data = )
```

ramka danych

```
> model.pusty=lm(zm.objasniana ~ 1)
```

Metoda eliminacji oparta na kryterium AIC:

```
> step(model.pelny, scope=list(upper=model.pelny,lower=model.pusty), direction="backward", k=2)
```

Metoda dołączania oparta na kryterium BIC:

```
> step(model.pusty, scope=list(upper=model.pelny,lower=model.pusty), direction="forward", k=log(length(zm.objasniana)))
```

6 Funkcje w pakiecie R

```
model.liniowy=lm(zmienna.objasniana~zmienna.objasnajaca1+zmienna.objasnajaca2)
```

- **wyraz wolny:** $\hat{\beta}_0$, gdzie $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$
`coef(model.liniowy)[1]`
`model.liniowy$coefficients[1]`
`model.liniowy$coef[1]`
- **współczynnik kierunkowy:** $\hat{\beta}_1$
`coef(model.liniowy)[2]`
- $\hat{\sigma}$, gdzie $\hat{\sigma}^2$ to nieobciążony estymator σ^2 : $\hat{\sigma}^2 = \frac{1}{n-(p+1)} \sum_{i=1}^n e_i^2$
`summary(model.liniowy)$sigma`
- **wartości prognozowane:** \hat{Y}_i
`fitted(model.liniowy)`
`model.liniowy$fitted.values`
`model.liniowy$fit`
- **residua (wartości resztowe):** $e_i = Y_i - \hat{Y}_i$
`residuals(model.liniowy)`
`resid(model.liniowy)`
`model.liniowy$residuals`
`model.liniowy$resid`
- **współczynnik determinacji:** $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$
`summary(model.liniowy)$r.squared`
- **skorygowany współczynnik determinacji:** $R_{adj}^2 = 1 - \frac{SSE/(n-(p+1))}{SST/(n-1)}$
`summary(model.liniowy)$adj.r.squared`
- **wpływy:** h_{ii}
`hatvalues(model.liniowy)`
- **residua standardyzowane:** $r_i = \frac{e_i}{SE_{e_i}}$, gdzie SE_{e_i} to esymator odchylenia standardowego

$$e_i \text{ równy } \hat{\sigma} \sqrt{1 - h_{ii}}$$
`rstandard(model.liniowy)`
- **studentyzowane residua modyfikowane:** $t_i = \frac{d_i}{SE_{d_i}}$, gdzie $d_i = Y_i - \hat{Y}_{i(i)}$
`rstudent(model.liniowy)`
- **odległości Cooke’a:** $D_i = \frac{1}{p+1} \cdot r_i^2 \cdot \frac{h_{ii}}{1-h_{ii}}$
`cooks.distance(model.liniowy)`
- **współczynniki podbicia wariancji:** VIF_i
`library(faraway)`
`vif(model.liniowy)`

$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$ - całkowita suma kwadratów (total sum of squares)

$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ - regresyjna suma kwadratów (sum of squares from the regression),

$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$ - suma kwadratów błędów (sum of squared errors)