

Podstawy analityki biznesowej

Ćwiczenie 3: Dane kategoryjne

Zadanie 1. Plik **racestalls.sas7bdat** zawiera numery przegród startowych zwycięzców w 144 wyścigach konnych zorganizowanych w USA. Wszystkie wyścigi miały miejsce na torze kołowym i wszystkie odnoszą się do wyścigów ośmiu koni. Przegroda oznaczona numerem 1 odpowiada torowi najbliższemu wnętrzu toru wyścigowego. Chcemy określić na ile szanse na wygraną zależą od pozycji startowej.

- (a) Zaimportować plik **racestalls.sas7bdat** do biblioteki **sasue** utworzonej na poprzednich zajęciach i wyświetlić jego zawartość. Zmienna **stall** oznacza numer przegród startowej zwycięzcy wyścigu.
- (b) W celu utworzenia szeregu rozdzielczego punktowego należy otworzyć **Zadania** ► **Statystyka** ► **Jednoczynnikowe liczebności**. W **Dane** ► **Dane** wybrać **sasue.racestalls**. W **Dane** ► **Role** ► **Zmienne analizowane** dodać **stall**, po czym kliknąć **Uruchom**. Czy procent zwycięzców z różnych przegród startowych znacząco się różni?
- (c) Sprawdzić, że ten sam wykres słupkowy można uzyskać wybierając **Zadania** ► **Wykresy** ► **Wykres słupkowy**. Alternatywnie można narysować wykres kołowy. Dokonuje się tego otwierając **Zadania** ► **Wykresy** ► **Wykres kołowy**, w **Dane** ► **Dane** wybierając **sasue.racestalls**, oraz w **Dane** ► **Role** ► **Zmienne kategoryzujące** dodając **stall**. Oczywiście, wszystko kończy kliknięcie **Uruchom**. Ten wykres również potwierdza, że liczby wygranych osiąganych z różnych przegród startowych różnią się.
- (d) Wykres słupkowy staje się często bardziej użyteczny jeśli słupki uporządkuje się w rosnącym lub malejącym porządku liczebności. Nie jest to jednak żadna z opcji dostępnych w zakładkach **Dane** lub **Opcje**. Można ją jednak uzyskać otwierając **Zadania** ► **Wykresy** ► **Wykres kołowy**, w **Dane** ► **Dane** wybierając **sasue.racestalls**, oraz w **Dane** ► **Role** ► **Zmienne kategoryzujące** dodając **stall**. W zakładce **Kod** odnajdujemy poniższy fragment:

```
proc sgplot data=SASUE.RACESTALLS;  
    /*--Bar chart settings--*/  
    vbar stall / name='Bar';  
    /*--Response Axis--*/  
    yaxis grid;  
  
run;
```

Dokonać edycji linii z **vbar** zamieniając ją na następującą:

```
vbar stall / name='Bar' categoryorder=respasc;
```

Teraz proszę uruchomić zmodyfikowany kod. Można go zapisać wybierając nad oknem edytora przycisk **Dodaj do Moje wstawki** oraz wpisując nazwę wstawki, np. **Uporządkowany wykres słupkowy**.

Odczytać z wykresu cztery przegrody startowe z największą liczbą zwycięzców i cztery przegrody z najniższą liczbą zwycięzców.

- (e) Gdyby pozycja przegrody startowej nie wpływała na szanse konia wygrywającego wyścig, oczekivalibyśmy że liczby zwycięstw z każdej przegrody są w przybliżeniu takie same. Hipotezą zerową byłaby więc identyczna liczba zwycięstw z każdej przegrody, równa $144/8=18$. Odpowiednim testem jest tu test zgodności chi-kwadrat, w którym statystyka testowa dla naszych konkretnych danych przyjmuje wartość (wydedukować jak się ją tworzy):

$$\frac{(29-18)^2}{18} + \frac{(19-18)^2}{18} + \frac{(18-18)^2}{18} + \frac{(25-18)^2}{18} + \frac{(17-18)^2}{18} + \frac{(10-18)^2}{18} + \frac{(15-18)^2}{18} + \frac{(11-18)^2}{18}$$

Jeśli hipoteza zerowa jest prawdziwa, rozważana statystyka ma rozkład chi-kwadrat o siedmiu stopniach swobody. Aby zastosować ten test, powróćmy do **Zadania ► Statystyka ► Jednoczynnikowe liczebności**. W **Opcje ► Statystyki ► Dobroć dopasowania chi-kwadrat** wybrać **Test asymptotyczny** i kliknąć **Uruchom**. W rezultatach odnaleźć odpowiednią *p*-wartość. Co mówi ona o pozycji przegrody startowej jako czynnika i szansach na wygraną konia? Jak zinterpretować wykres odchyleń względnych?

Zadanie 2. Plik **self_exam.sas7bdat** zawiera rezultaty ankiety, w której pytano kobiety jak często dokonują profilaktycznej samokontroli piersi. Zaimportować go do biblioteki **sasue** i wyświetlić zawarte w nim dane. W ćwiczeniu chcemy zbadać czy częstość samokontroli zależy od wieku kobiety.

- (a) Odpowiedni test chi-kwadrat rozpoczyna się od otworzenia **Zadania ► Statystyka ► Analiza kontyngencji**. W **Dane ► Dane** wybiera się **sasue.self_exam**. W **Dane ► Role ► Zmienne wierszowe** dodać **agegroup**, a w **Dane ► Role ► Zmienne kolumnowe** dodać **frequency**. Rozwinąć **Dodatkowe role** i przypisać **num** jako **Liczba wystąpień**. W **Opcje ► Statystyki** zauważyć, że **Statystyki chi-kwadrat** jest wybrane domyślnie. W **Opcje ► Wykresy** wybrać **Zablokuj wykresy** w celu uproszczenia rezultatów. Kliknąć **Uruchom**. Odszukać wartość statystyki chi-kwadrat i odpowiadającą jej *p*-wartość. Co wynika z tej ostatniej (hipotezą zerową była niezależność wieku i częstości samokontroli)?
- (b) Tabelę dwudzielczą (to ta z nagłówkiem **Tabela agegroup od frequency**) oraz test chi-kwadrat często uzupełnia się podając wartość znaną jako **iloraz szans**. Jest ona miarą siły związku między dwoma zmiennymi binarnymi, dla których buduje się tabelę. Aby szczegółowiej wyjaśnić to pojęcie, zauważmy że prawdopodobieństwo tego, że młodsza kobieta rzadko samokontroluje piersi można ocenić jako $141/232 = 0.61$, a prawdopodobieństwo, że samokontroluje piersi co miesiąc, można oszacować jako $91/232 = 0.39$, czyli $1 - 0.61$. Iloraz tych dwóch ocen prawdopodobieństwa daje stosunek szansy rzadkiej samokontroli do szansy comiesięcznej samokontroli. Tu jego wartość wynosi 1.54, czyli więcej od jedności, co wskazuje, że proporcja młodszych kobiet rzadko samokontrolujących piersi jest większa od proporcji młodych kobiet samokontrolujących piersi raz w miesiącu.

Dla kobiet w wieku 45 lat i starszych, odpowiednie oszacowanie stosunku szans wynosi 2.72, co ponownie wskazuje, że dla starszych kobiet proporcja dokonująca regularnej samokontroli jest mniejsza od proporcji dokonującej samokontroli okazjonalnie.

Iloraz dwóch powyższych stosunków szans jest oceną rozważanego ilorazu szans, która wynosi $1.54/2.72 = 0.6$. Iloraz szans równy jedności odpowiadałby niezależności zmiennych binarnych. Opisany wcześniej test chi-kwadrat jest więc zasadniczo testem sprawdzającym, czy iloraz szans ma wartość jeden. Dla ilorazu szans można jednak dodatkowo zbudować przedział ufności, który pozwala powiedzieć coś więcej o sile związku między dwoma rozważanymi zmiennymi.

Proszę powtórzyć czynności z punktu (a), przed kliknięciem **Uruchom** wybierając **Iloraz szans i ryzyka względne (dla tabel 2x2)** w **Opcje ► Statystyki**.

Proszę zauważyć, że wśród rezultatów pojawiła się nowa tabela z nagłówkiem **Iloraz szans i ryzyka względne**. Jaki jest 95% przedział ufności dla ilorazu szans? Czy zawiera on wartość jeden? Co to oznacza i czy zgadza się to z wynikiem testu chi-kwadrat?

Zadanie 3. Plik **ca_lung.sas7bdat** zawiera dane zebrane w 1992 r., mające stanowić podstawę do zbadania zależności między paleniem papierosów i rakiem płuc w czterech miastach chińskich (Pekin, Szanghaj, Nanjing i Zhengzhou). Zaimportować go do biblioteki **sasue** i wyświetlić zawarte w nim dane.

- (a) Kuszącym jest podejście polegające na połączeniu rezultatów z różnych miast w całość, przez co uzyskalibyśmy tylko dwie zmienne binarne (**smoker** oraz **cancer**, z wyeliminowaną zmienną **city**) i możliwość wykorzystania podejścia opisanego w poprzednich zadaniach. Jest to jednak niebezpieczne o tyle, że mogłoby doprowadzić do wygenerowania relacji nie istniejących w rzeczywistości lub zamaskowania prawdziwych związków między zmiennymi. Analizę uzależniamy więc od miasta, a jej narzędziem będzie tzw. metoda Cochran-Mantel-Haenszela (CMH).

Proszę otworzyć **Zadania ► Statystyka ► Analiza kontyngencji**. W **Dane ► Dane** wybiera się **sasue.ca_lung**. W **Dane ► Role ► Zmienne wierszowe** dodać **smoker**, w **Dane ► Role ► Zmienne kolumnowe** dodać **cancer**, a w **Dane ► Role ► Zmienne warstwowe** dodać **city**. Rozwinąć **Dodatkowe role** i przypisać **count** jako **Liczba wystąpień**. W **Opcje ► Statystyki** wybrać **Statystyki Cochran-Mantel-Haenszela**. W **Opcje ► Wykresy** wybrać **Zablokuj wykresy** w celu uproszczenia rezultatów. Kliknąć **Uruchom**.

- (b) W tabeli **Test Breslowa-Daya na homogeniczność ilorazów szans** odszukać wartość statystyki chi-kwadrat, która ma trzy stopnie swobody (o jeden mniej od liczby tabel) i odpowiadającą jej *p*-wartość. Co wynika z tej ostatniej (hipotezą zerową była niezależność ilorazu szans dla związku między paleniem i nowotworem od miasta)?
- (c) Ponieważ nie mamy istotnych dowodów na odrzucenie hipotezy o niezależności wyników od miasta, możemy połączyć dane z różnych miast w całość i użyć ich do oceny wspólnego ilorazu szans dla wszystkich miast. Odnaleźć tę ocenę wraz z jej przedziałem ufności w tabeli **Wspólny iloraz szans i ryzyka względne**. Czy występowanie nowotworu płuc zależy od palenia, a jeśli tak – to jak mocno?

Zadanie 4. Plik **tumors.sas7bdat** zawiera dane dotyczące badania 141 osób chorujących na nowotwór mózgu. Rozróżnia się tu trzy rodzaje nowotworu: A – nowotwór łagodny, B – nowotwór złośliwy, oraz C – inne nowotwory mózgu. Oprócz tego rozróżnia się trzy miejsca występowania nowotworu: I – płaty czołowe, II – płaty skroniowe, III – inne obszary mózgu. Zaimportować plik do biblioteki **sasue** i wyświetlić zawarte w nim dane. Jesteśmy zainteresowani ewentualnymi

dowodami na to, że pewne typy nowotworu występują częściej w szczególnych obszarach mózgu (czyli tym, czy istnieje związek między zmiennymi kategorialnymi **type** oraz **site**).

- (a) Rozważane dane można przedstawić w postaci tablicy kontyngencji o rozmiarze 3×3 . W tym celu należy otworzyć **Zadania ▶ Statystyka ▶ Analiza kontyngencji**. W **Dane ▶ Dane** wybiera się **sasue.tumors**. W **Dane ▶ Role ▶ Zmienne wierszowe** dodać **site**, w **Dane ▶ Role ▶ Zmienne kolumnowe** dodać **type**. W **Opcje ▶ Statystyki** zauważyć, że domyślnie wybiera się **Statystyki chi-kwadrat**. W **Opcje ▶ Wykresy** wybrać **Zablokuj wykresy** w celu uproszczenia rezultatów. Kliknąć **Uruchom**. Tabela kontyngencji jest pierwszą z wyświetlonych tabel.
- (b) Hipotezą zerową jest niezależność miejsca i typu nowotworu. Oznacza to tyle, że prawdopodobieństwa typów nowotworu są takie same we wszystkich miejscach. Ściślej, niezależność implikuje, iż prawdopodobieństwo tego, że pacjent ma nowotwór określonego typu w określonym miejscu jest iloczynem prawdopodobieństwa wystąpienia danego typu nowotworu i prawdopodobieństwa wystąpienia nowotworu w danym miejscu. Oba ostatnie prawdopodobieństwa można oszacować dzieląc sumy obserwacji w odpowiedniej kolumnie i odpowiednim wierszu w tablicy kontyngencji przez całkowitą liczbę obserwacji. Przykładowo, oceną prawdopodobieństwa nowotworu A jest $78/141 = 0.553$, oceną prawdopodobieństwa nowotworu w miejscu I natomiast $38/141 = 0.270$. Jeśli więc hipoteza zerowa jest prawdziwa, oceną prawdopodobieństwa, że pacjent ma nowotwór A w miejscu I wynosi $0.553 \times 0.270 = 0.149$. Wyniak stąd, że przy założeniu niezależności oczekiwana liczba pacjentów z nowotworem A w miejscu I wynosi $141 \times 0.149 = 21.0$. W ten sam sposób można wyznaczyć wartości oczekiwane w pozostałych komórkach tablicy kontyngencji. Mogą one być później porównane z zaobserwowanymi wartościami w oparciu o test chi-kwadrat. Dla tablicy kontyngencji o r wierszach i c kolumnach, test chi-kwadrat ma $(r-1)(c-1)$ stopni swobody (w rozważanym przykładzie będą to więc 4 stopnie swobody).

Odszukać wartość statystyki chi-kwadrat i skojarzoną z nią p -wartość. Z tej drugiej wynika, że nie ma bardzo silnego dowodu przeciw hipotezie o niezależności typu i miejsca nowotworu.

- (c) Można spróbować określić komórki tabeli kontyngencji, dla których odstępstwa od prawdziwości hipotezy zerowej są największe. W tym celu można porównywać liczebności zaobserwowane i oczekiwane przy założeniu prawdziwości hipotezy o niezależności – różnice między tymi wartościami nazywa się *resztami*. Niestety, taka prosta definicja reszt nie jest satysfakcjonująca, ponieważ ich duże wartości często występują dla komórek mających również duże oczekiwane wartości liczebności, co może być mylące. Bardziej sensowną definicją residuum byłoby

$$\frac{\text{liczebność zaobserwowana} - \text{szacowana liczebność oczekiwana}}{\sqrt{\text{szacowana liczebność oczekiwana}}}$$

co prowadzi do tzw. *reszt Pearsona*. Może to jednak prowadzić do zbyt konserwatywnych rezultatów (mogą być wskazywane również komórki, które sensownie spełniają założenie o niezależności). W praktyce zaleca się więc używanie następującej definicji reszty:

$$\frac{\text{reszta Pearsona}}{\sqrt{\left(1 - \frac{\text{suma w wierszu}}{\text{rozmiar próby}}\right) \left(1 - \frac{\text{suma w kolumnie}}{\text{rozmiar próby}}\right)}}$$

nazywanej *resztą dopasowaną* lub *resztą standaryzowaną*.

Pierwsza z trzech typów reszt jest dostępna w **Opcje ► Tabela liczebności ► Liczebności** po wybraniu **Odchylenie**. Pozostałe wymagają edycji kodu, czego dokonuje się następująco:

W **Opcje ► Tabela liczebności ► Liczebności** wybieramy **Obserwowane**, **Oczekiwane** oraz **Odchylenie**. W panelu kodu, po wyborze Edytuj, kod powinien wyglądać następująco:

```
proc freq data=SASUE.TUMORS;  
    tables (site) *(type) / chisq expected deviation nopercnt  
norow nocol nocum  
    plots=none;  
run;
```

Po `plots = none`, ale przed średnikiem, należy wpisać `crosslist(pearsonresstdres)`.

Dla jakich komórek tablicy kontyngencji występują dwa największe odchylenia od hipotezy o niezależności, wyrażone w postaci reszt standaryzowanych? Co oznaczają znaki tych reszt?

Zadanie 5. W psychologii bada się zachowania tłumów obecnych przy próbach samobójczych. Plik **baiting.sas7bdat** zawiera dane dotyczące 21 przypadków próby skoku z dachu budynku i zachowania tłumy podlegającego do skoku (*ang. baiting*) lub zachowania drwiącego z samobójcy (*ang. jeering*). Dane zawierają również informację o porze roku (*warm* oznacza okres od czerwca do września, a *cool* – okres od października do maja). Hipoteza postawiona przez Leona Manna w 1981 r. stwierdzała, że podleganie jest bardziej prawdopodobne w porze cieplej. Kuszającym jest zastosowanie tu poznanego już testu chi-kwadrat, jednak nie powinno się go używać z uwagi na to, że oczekiwane liczebności są zbyt małe (tzn. mniejsze od pięciu). W jego miejsce w przypadku tablic kontyngencji o rozmiarze 2×2 przyjęło się stosować tzw. dokładny test Fishera.

- (a) Skopiować plik **baiting** do biblioteki **sasue**, a następnie wyświetlić jego zawartość.
- (b) Otworzyć **Zadania ► Statystyka ► Analiza kontyngencji**. W **Dane ► Dane** wybiera się **sasue.baiting**. W **Dane ► Role ► Zmienne wierszowe** dodać **season**, w **Dane ► Role ► Zmienne kolumnowe** dodać **baiting**. W **Dane ► Dodatkowe role ► Liczba wystąpień** dodać **count**. W **Opcje ► Statystyki** zauważyć, że domyślnie wybiera się **Statystyki chi-kwadrat**, ale wybór **Test dokładny ► Test dokładny Fishera** nie został zaznaczony. Można byłoby go wybrać, jednak dla tablic 2×2 jest on stosowany domyślnie. W **Opcje ► Wykresy** wybrać **Zablokuj wykresy** w celu uproszczenia rezultatów. Kliknąć **Uruchom**.
- (c) W tabeli **Test dokładny Fishera** odnaleźć odpowiednią *p*-wartość. Czy daje ona mocny dowód do odrzucenia hipotezy o niezależności zachowania tłumy i pory roku? Proszę uwzględnić, że wielkość próby jest mała i test ma raczej małą moc. Jak zachowuje się w tej sytuacji test chi-kwadrat?

Zadanie 6. Dokładny test Fishera można stosować również do większych tablic kontyngencji. Plik **lesions.sas7bdat** zawiera dane dotyczące liczby wystąpień (zmienna **n**) nowotworu jamy ustnej w trzech wiejskich regionach geograficznych Indii (zmienna **region**) wraz z informacją o miejscu wystąpienia lezji w jamie ustnej (zmienna **site**). Z uwagi na małe liczebności, adekwatny jest tu właśnie test dokładny.

- (a) Otworzyć **Zadania ► Statystyka ► Analiza kontyngencji**. W **Dane ► Dane** wybiera się **sasue.lesions**. W **Dane ► Role ► Zmienne wierszowe** dodać **site**, w **Dane ► Role ► Zmienne kolumnowe** dodać **region**. W **Dane ► Dodatkowe role ► Liczba wystąpień** dodać **n**. W **Opcje ► Statystyki** zauważyć, że domyślnie wybiera się **Statystyki chi-kwadrat**, ale tym razem zaznaczyć wybór **Test dokładny ► Test dokładny Fishera**. W **Opcje ► Wykresy** wybrać **Zablokuj wykresy** w celu uproszczenia rezultatów. Kliknąć **Uruchom**.
- (b) W tabeli **Test dokładny Fishera** odnaleźć odpowiednią p -wartość. Co mówi ona o niezależności regionu geograficznego i miejscu wystąpienia nowotworu? Jak zachowuje się w tej sytuacji test chi-kwadrat? Który test przekłamuje?

Zadanie 7. Plik **arrests.sas7bdat** zawiera dane dotyczące próby nieletnich skazanych w 1987 r. na Florydzie za popełnienie przestępstwa. Przestępców dobrano w pary na podstawie kryteriów takich jak wiek i liczba poprzednio popełnionych przestępstw. Dla każdej z par, jedną osobą zajmował się są dla nieletnich, a drugą – sąd dla dorosłych. W badaniu odnotowano również czy młodociani przestępcy byli ponownie aresztowani przed końcem 1988 r. Przedmiotem badania jest określenie, czy proporcje ponownie aresztowanych są identyczne dla sądów dla dorosłych i dla nieletnich.

- (a) Wyświetlić zawartość pliku **arrests**. Test chi-kwadrat zakłada niezależność obserwacji. Dane w rozważanym pliku nie są jednak niezależne, bo pochodzą z dopasowanych par. Przykładowo, dla 158 par przestępców zmienne **adult** i **junevile** są równe **y** (yes), co oznacza, że obydwaj przestępcy w każdej z tych par zostali ponownie aresztowani. Aby przetestować, czy wskaźnik ponownego aresztowania różni się między sądami dla dorosłych i nieletnich, zastosujemy tzw. *test McNemara*.
- (b) Otworzyć **Zadania ► Statystyka ► Analiza kontyngencji**. W **Dane ► Dane** wybiera się **sasue.arrests**. W **Dane ► Role ► Zmienne wierszowe** dodać **adult**, w **Dane ► Role ► Zmienne kolumnowe** dodać **juvenile**. W **Dane ► Dodatkowe role ► Liczba wystąpień** dodać **num**. W **Opcje ► Statystyki** usunąć zaznaczenie przy **Statystyki chi-kwadrat** i zaznaczyć **Miary zgodności**. W **Opcje ► Wykresy** wybrać **Zablokuj wykresy** w celu uproszczenia rezultatów. Kliknąć **Uruchom**.
- (c) W tabeli **Test McNemara** odnaleźć odpowiednią p -wartość. Co mówi ona o niezależności typu sądu i prawdopodobieństwa ponownego aresztowania? Przed jakim sądem rozprawa daje mniejsze szanse na ponowne aresztowanie?

Zadanie 8. Plik **crash** zawiera liczebności śmiertelnych wypadków powietrznych w Australii z podziałem na kwartały roku, zebrane w ciągu 20 kolejnych lat. Ocenic hipotezę, że wskaźniki wypadków są jednakowe w przekroju tych czterech kwartałów.

Zadanie 9. Stu Amerykanów poproszono o wybranie spośród pięciu pozycji tej, której najbardziej się obawiają. Wyniki zawiera plik **fear**. Przetestować czy płęć i największa obawa są niezależne od siebie.

Zadanie 10. Pacjentów psychiatrycznych można sklasyfikować jako psychotycznych lub neurotycznych. Plik **suicidal** zawiera wyniki badania 20 pacjentów, w którym rejestrowano również, czy pacjenci mieli myśli samobójcze. Jakie wnioski można wysnuć o zależności między typem pacjenta a występowaniem myśli samobójczych?

Zadanie 11. Plik **cancer** zawiera dane z badania częstości używania doustnych środków antykoncepcyjnych wśród 183 kobiet chorujących na nowotwór endometrium. Każdej kobiecie dopasowano odpowiedniczkę (wg kryterium wieku, szpitala, dat zażywania środków itp.), która nie

zachorowała na nowotwór. Przykładowo, były 43 pary, wśród których kobieta chorująca na nowotwór zażywała środki antykoncepcyjne (zmienna **caseuse** równa **Y**), a kobieta nie chorująca na nowotwór nie zażywała środków antykoncepcyjnych (zmienna **controluse** równa **N**). Czy dane dostarczają dowodu na istnienie zależności między zażywaniem środków antykoncepcyjnych i nowotworem endometrium?

Zadanie 12. Plik **infdeath** zawiera dane dotyczące wieku śmierci dzieci, które zmarły z powodu dziewięciu różnych przyczyn. Zmienna **age** ma wartości **1** (0 dni), **2** (1–6 dni), **3** (2–4 tygodnie), **4** (5–26 tygodni) oraz **5** (27–51 tygodni). Z badać zależność między przyczyną śmierci i wiekiem dziecka.

Zadanie 13. W latach 80. w USA przeprowadzono badanie dotyczące obwiniania ofiary w procesach dotyczących gwałtów. Dane przedstawia poniższa tabela:

	werdykt skazujący		
	tak	nie	razem
nisko	153	24	177
wysoko	105	76	181
razem	258	100	358

Wartości **nisko** i **wysoko** odpowiadają temu, jak mocno obrona oskarżała ofiarę o prowokowanie do domniemanego gwałtu.

Czy powyższe dane dają powód do stwierdzenia, że sposób prowadzenia obrony oskarżonego ma wpływ na końcowy werdykt sądu?