

## Statystyka II. Egzamin komputerowy 04.09.2007.

1. Niech czas życia żarówki ( $C$ ) ma rozkład gamma z parametrami: skala=1, kształt=3. Wylosuj 1000 niezależnych obserwacji z tego rozkładu i oszacuj prawdopodobieństwo, że  $C$  jest nie większy niż  $d=1.5 \cdot z_{0.75}$ , gdzie  $z_{0.75}$  jest kwantylem próbkowym rzędu 0.75 tego rozkładu (wskazówka: funkcja *quantile* oblicza kwantyle próbkowe). Policz  $d$  numerycznie (bez losowania).
2. Zbuduj model analizy wariancji dla danych `immer` z biblioteki `MASS`. W modelu tym zmienną objaśnianą jest  $Y = (Y1+Y2) / 2$  - średni plon jęczmienia z lat 1931-32. Dobierz transformację zmiennej  $Y$  korzystając z funkcji `boxcox` i `logtrans` oraz wykonaj automatyczną selekcję zmiennych.
3. Na danych demograficznych o szkołach holenderskich `MathAchSchool[, -c(1, 6)]` z biblioteki `nlme`, porównaj w eksperymencie weryfikacji krzyżowej (4/5 danych do trenowania) metody klasyfikacji zaimplementowane w funkcjach `rpart`, `nnet` oraz regresję logistyczną. Cecha przewidywana przyjmuje dwie wartości:  $y = \text{ifelse}(X[, 10] == \text{"Catholic"}, 1, 0)$ . Policz współczynnik Goodmana-Kruskala między  $y$  oraz  $y\_pred$  dla tych trzech metod.
4. Wczytaj z archiwum `MacierzePot12_3kol.zip` potencjały kontaktowe (PK), czyli 12 symetrycznych macierzy  $20 \times 20$  opisujących oddziaływania 210 par aminokwasów w białkach. PK są zapisane w plikach trójkolumnowych tak, że w pierwszej kolumnie znajduje się numer wiersza, w drugiej – numer kolumny, a w trzeciej wartość odpowiedniego elementu macierzy PK o tych współrzędnych. Napisz ogólną funkcję `three2Vec`, która przekształca plik trójkolumnowy na wektor złożony z elementów macierzy PK należących do jej górnej części (ewentualnie dolnej) oraz przekątnej. Ze zbioru wektorów  $x_1, \dots, x_{12}$  otrzymanych z 12 PK, zbuduj macierz danych  $X = [x_1, \dots, x_{12}]$ . Sprawdź, czy w zbiorze  $X$  -- 210 obserwacji 12-wymiarowych są obserwacje odstające.
5. Niech  $p = (p_1, \dots, p_k)$  będzie wektorem prawdopodobieństw. Zaimplementuj funkcję rysującą wykres konturowy 2-wymiarowej gęstości  $f(x, y)$  tak, że warstwice ograniczają obszary o zadanych prawdopodobieństwach  $p = (p_1, \dots, p_k)$ . Na przykład, jeśli  $p = c(1, 2, 3, 4) / 5$ , to pierwsza warstwica jest zadana przez zbiór  $\{(x, y) : f(x, y) = z, P(f(x, y) \geq z) = 1/5\}$  dla pewnego  $z$ . Policz gęstość (`PRACAD`, `DISCLIM`) za pomocą `kde2d` i narysuj warstwice dla danego wyżej  $p$  na danych o szkołach holenderskich.