EGZAMIN ZE STATYSTYKI II - 27.07.2005

1. Niech V_4 będzie modyfikacją współczynnika Ginniego $V_4(p) = 1 - \sum_{i=1}^{n} p_i^4$, gdzie $p = (p_1, \ldots, p_n)$ jest gęstością prawdopodobieństwa skoncentrową na $\{1, \ldots, n\}$. Udowodnij, że rozkład jednostajny $q=(1/n,\ldots,1/n)$ maksymalizuje V_4.

2. Przy pomocy V_4 definiujemy modyfikację współczynnika Goodmana-Kruskala

   $\tau_4(Y/X) = \dfrac{V_4(Y) - EV_4(Y/X)}{V_4(Y)}$. Udowodnij, że $0 \le \tau_4 \le 1$.

3. Podaj algorytm klasteryzacji aglomeracyjnej „average linkage”. W jaki sposób dendrogram indukuje podział zbioru obserwacji na k-części ?

4. Niech $X = (X_1, \ldots, X_p)^T$ będzie wektorem losowym. Dla jakiego wektora współczynników $t = (t_1, \ldots, t_p)^T$ takiego, że $t^T t = 1$, $\mathrm{var}(t^T X)$ jest maksymalna ? Ile wynosi to maksimum ?

5. Niech $Y = X\beta + \varepsilon$ będzie modelem regresji wielorakiej oraz $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$ estymatorem najmniejszych kwadratów $\beta = (\beta_1, \ldots, \beta_p)^T$. Udowodnij, że dla $i=1,\ldots,n$ $\hat{\beta}_i$ jest nieobciążonym estymatorem β_i o minimalnej wariancji wśród estymatorów liniowych.