Statystyka I semestr zimowy 2017, seria XII

1. **Analiza wariancji (ANOVA)** Załóżmy, że mamy k różnych grup obserwacji, a w każdej grupie $j = 1, \ldots, k$ mamy niezależne obserwacje $Y_{1,j}, \ldots, Y_{n_j,j} \sim \mathcal{N}(\mu_j, \sigma^2)$. Naszym celem jest testowanie hipotezy $H_0: \ \mu_1 = \mu_2 = \cdots = \mu_k$ przeciw $H_1: \mu_i \neq \mu_j$ dla pewnych $i \neq j$. Zapisz powyższy problem w postaci modelu liniowego i podaj odpowiedni test ilorazu wiarygodności.

2. (a) Pokaż, że test z poprzedniego zadania jest równoważny testowi postaci $F = \frac{BSS/(k-1)}{WSS/(n-k)} > c$, gdzie $WSS = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{i,j} - \bar{Y}_j)^2$, $BSS = \sum_{j=1}^{k} n_j (\bar{Y}_j - \bar{Y})^2$ oraz $n = n_1 + \cdots + n_k$.

   (b) Pokaż, że statystyka F ma rozkład F-Snedecora $F(k-1, n-k)$.

3. Rozważmy model liniowy $Y = X\beta + \epsilon$, w którym X jest macierzą $n \times p$ pełnego rzędu p, przy czym pierwsza kolumna X jest równa $\mathbf{1}$ oraz $\epsilon \sim \mathcal{N}(0, \sigma^2 Id)$. Niech dany będzie *współczynnik dopasowania* modelu do danych $R^2 = 1 - SSE/SST$, gdzie $SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ oraz $SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$. Pokaż, że

$$R^2 = \left(\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}} \right)^2 .$$

4. (a) Niech R^2 będzie takie, jak w poprzednim zadaniu. Pokaż, że test $R^2 > c$ jest równoważny testowi ilorazu wiarygodności dla modelu liniowego *zerowego* tj. modelu z macierzą X ograniczoną tylko do pierwszej kolumny.

   (b) Pokaż, że współczynnik dopasowania R^2 ma rozkład $Beta(\frac{p-1}{2}, \frac{n-p}{2})$.

5. Podaj estymatory największej wiarygodności parametrów μ oraz Σ na podstawie próby prostej $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \Sigma)$, gdzie Σ jest macierzą odwracalną wymiaru $d \times d$.