PSM Fall 2015, Exercise Sheet 8

Return on Tuesday Nov 10 in class

Note. You are encouraged to work in teams of two – but no larger. If you work in a team, submit only a single sheet with both names indicated on it. Nicely type-set solutions are highly appreciated.

Problem 1 (20 points total) Let $(X_1, \dots, X_n)$ be independent RVs with sample space $(\mathbb{R}, \mathcal{B})$ and statistical model $\mathcal{P} = \{P_\theta^X : \theta \in \Theta\}$ with unknown means $E_\theta[X_i] = c_i\mu$, $\mu \in \mathbb{R}$, unknown variance $VAR_\theta[X_i] = h_i\sigma^2$, $\sigma^2 \in (0, \infty)$ and known factors $c_i, h_i \in \mathbb{R}$.

Consider quadratic loss $L(\theta, d) = (\theta - d)^2$ and the estimator for the unknown mean parameter μ given by

$$\hat{\mu} = \frac{1}{\sum\limits_{i=1}^{n} \frac{c_i^2}{h_i}} \sum_{i=1}^{n} \frac{c_i}{h_i} X_i.$$

a. Show that $\hat{\mu}$ is best linear unbiased estimator (BLUE) for μ.

b. Show that $VAR[\hat{\mu}] = \frac{\sigma^2}{\sum\limits_{i=1}^{n} \frac{c_i^2}{h_i}}$

Problem 2 (30 points total) Let (X_1, X_2) be independent RVs with sample space $(\mathbb{R}, \mathcal{B})$ and statistical model $\mathcal{P} = \{P_\theta^X : \theta \in \Theta\}$ with unknown means $E_\theta[X_1] = \mu = E_\theta[X_2]$, $\mu \in \mathbb{R}$, variances $VAR_\theta[X_1] = 1$, $VAR_\theta[X_2] = \sigma^2$, for unknown $\sigma^2 \in (0, \infty)$.

Let $\mathcal{L} = \{T_a : T_a = aX_1 + (1-a)X_2, a \in \mathbb{R}\}$ be the class of linear and unbiased estimators in μ. Show that

a. (10 points) procedure T_0 is better than procedure T_a for all $a < 0$;

b. (10 points) procedure T_1 is better than procedure T_a for all $a > 1$;

c. (10 points) there exists no BLUE for μ for $0 \leq a \leq 1$.

Problem 3 (20 points total) Consider the simple linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad \forall i = 1, \dots, n,$$

for a dependent variable Y and one explanatory variable X, where ϵ_i are independent errors with mean $E[\epsilon_i] = 0$ and variance $VAR[\epsilon_i] = \sigma^2$.

a. (10 points) Compute the BLUE $\hat{\beta}$ for the slope parameter β in this model.

b. (10 points) Compute the BLUE $\hat{\alpha}$ for the intercept α in this model.

Problem 4 (30 points total) The data set `animalslogw.txt` contains the logarithmic measurements of body and brain weight for a selection of animals.

a. (10 points) Compute the linear regression equation modeling logarithmic brain weight as a function of logarithmic body weight using the least-squares approach.

b. (10 points) For the slope parameter β, test the hypothesis $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$ at the 5%-significance level.

c. (10 points) Compute the leverage of each data point (i.e. the diagonal elements of the hat matrix). Are there any observations with a high leverage effect?