Si consideri il dataset `longley` presente nella libreria `datasets`. La variabile risposta è `Employed`, i predittori sono `GNP.deflator`, `GNP`, `Unemployed`, `Armed.Forces`, `Population` e `Year`.

a) Per questi dati, si stimi la regressione *Best Subset Selection* scegliendo il *Best Subset* finale con il criterio BIC. Riportare la stima $\hat{\beta}_{`GNP`}$ per la variabile `GNP`.

b) Si suddivida il dataset in *Learning set* con osservazioni con indici in $L = \{1, 3, 5, 7, 9, 11, 13, 15\}$ e *Inference set* con osservazioni con indici in $I = \{2, 4, 6, 8, 10, 12, 14, 16\}$. Sulla base del *Learning set*, stimare $\hat{S} = \{j : \hat{\beta}_j \neq 0\}$, dove $\hat{\beta}_j$ sono le stime della regressione *Best Subset Selection* scegliendo

il *Best Subset* finale con il criterio BIC. In altre parole, $\hat{S}$ contiene le variabili selezionate da *Best Subset Selection* stimato sul *Learning set*. Sulla base dell'*Inference set*, calcolare i p-values del modello lineare con le variabili selezionate (e l'intercetta). Riportare il p-value relativo alla variabile `Unemployed` aggiustato con il metodo di Bonferroni, che tiene conto della molteplicità della selezione.

c) Sia $\hat{\mu}_L(x) = \hat{\mu}(x; (x_l, y_l), l \in L)$ il modello *Best Subset Selection* stimato sul *Learning set* al punto precedente. Calcolare i residui in valore assoluto $R_i = |y_i - \hat{\mu}_L(x_i)|$ per $i \in I$, e ordinare $\{R_i, i \in I\}$ in senso crescente, i.e. $R_{(1)} \leq \ldots \leq R_{(m)}$. Riportare il valore critico $R_\alpha = R_{(k)}$ con $k = \lceil (1 - \alpha)(m + 1) \rceil$, dove $m = 8$ e $\alpha = 1/3$.