|  |  |
| --- | --- |
| ESAME DI MODELLI STATISTICI | |
| Prof. Antonio Lucadamo | 22 luglio 2021 |

|  |  |  |  |
| --- | --- | --- | --- |
| Candidato |  | Matricola |  |

|  |
| --- |
| Esercizio 1 |
| L’insieme di dati Pima.tr disponibile nella libreria MASS di R contiene informazioni su un campione di donne di almeno 21 anni che appartengono alla riserva indiana Pima, nei pressi di Phoenix, Arizona. Le variabili osservate sono: numero di gravidanze (npreg), concentrazione di glucosio (glu), pressione sanguigna (bp), spessore della pelle del tricipite in millimetri (skin), indice di massa corporea (bmi), età in anni (age) e presenza/assenza di diabete (type).   1. Specificare e stimare un modello appropriato per studiare in che modo la probabilità di contrarre il diabete dipende dalle altre variabili. 2. Dopo aver stimato il modello che include tutte le variabili esplicative in esame, verificare l’ipotesi nulla che i coefficienti associati alle variabili npreg, bp, skin siano contemporaneamente nulli. 3. Commentare il modello ridotto stimato in termini di rapporto delle quote 4. Costruire un intervallo di confidenza al 95% per il coefficiente della variabile bmi. Confrontare i risultati corrispondenti all’intervallo di tipo Wald e a quello basato sulla log-verosimiglianza profilo. 5. Costruire un intervallo di confidenza al 95% per la probabilità di riscontrare il diabete quando: glu=190, bmi=35, ped=0.80, age=45 6. Valutare sensitività e specificità usando come valore soglia c = 0.5 7. Calcolare l’area sotto la curva ROC, descrivendo brevemente la procedura adottata |

|  |
| --- |
| Esercizio 2 |
| Caricare il dataset fish con il seguente comando:  fish<-read.csv(<https://stats.idre.ucla.edu/stat/data/fish.csv>)   1. Specificare un modello opportuno per studiare in che modo la variabile count dipenda dalle variabili camper, child e persons. Non fare distinzione fra le variabili incluse nelle due parti del modello 2. Riutilizzare lo stesso modello individuato al punto precedente, utilizzando la variabile persons per modellare la “parte” binomiale e le altre due per la restante parte del modello. 3. Valutare l’utilizzo di un ulteriore modello che ci consenta di studiare tale tipologia di dati e definire quale fra i modelli può essere ritenuto più opportuno. 4. Interpretare i risultati ottenuti, concentrandosi sul modello migliore e descrivendo eventuali differenze con gli altri. 5. Utilizzare il modello finale per effettuare una previsione per la variabile count quando child=2, camper=1 e persons=4. |