**EXAM**

**STATISTICAL METHODS FOR TERRITORIAL AND SOCIAL ENVIRONMENTS**

1. **The multiple regression model and its goodness of fit.**

In 2021 was published a paper about multiple regression analysis to forecast tourist arrivals on the Island of Boracay, Malay, Aklan, Philippines. The research sought to identify and analyse the factors influencing these arrivals to aid in effective tourism management and planning:

* *Y -* is travel and tourism total contribution to GDP (in billion Euro),
* *X*1 - Amenities: Score given by the tourists on Facilities and services available to tourists,
* *X*2 - Accommodation: Score given by the tourists on Lodging and places where tourists can stay,
* *X*3 - Accessibility: Score given by the tourists on Ease of access to the island, including transportation options and infrastructural elements,
* *X*4 - Attraction: Score given by the tourists on Tourist spots, natural beauty, activities, and overall appeal of the island.

The authors provided the results for the multiple regression model, reported in Table 1.

| **Variable** | | **Coefficients** | **Standard Error** | | **t Stat** | | **P-value** | **Lower 95%** | | **Upper 95%** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Intercept** | | -1,025,249.6 | 2,287.4306 | | -448.21014 | | 0.0014204 | -1,054,314.1 | | -996,185.03 | |
| **Accommodation** | | -223,261.47 | 9,906.7083 | | -22.536392 | | 0.02823 | -349,138.13 | | -97,384.802 | |
| **Attraction** | | 157,768.78 | 3,639.837 | | 43.345012 | | 0.0146847 | 111,520.26 | | 204,017.29 | |
| **Accessibility** | | 251,969.82 | 6,095.0339 | | 41.340184 | | 0.0153965 | 174,525.07 | | 329,414.57 | |
| **Amenities** | | 69,730.629 | 996.31802 | | 69.988325 | | 0.0090955 | 57,071.209 | | 82,390.05 | |
|  |
| **Component** | **df (Degrees of Freedom)** | | | **SS (Sum of Squares)** | | **MS (Mean Square)** | | | **F (F-statistic)** | | **Significance F** | |
| Regression | 4 | | | 6,785,342,362 | | 1,696,335,591 | | | 119,018.2163 | | 0.002173971 | |
| Residual | 1 | | | 14,252.7391 | | 14,252.73914 | | |  | |  | |
| Total | 5 | | | 6,785,356,615 | |  | | |  | |  | |

*Table 1: results for the multiple regression model.*

1. Write the estimated multiple linear regression model and comment the estimated parameters.
2. How well do the explanatory variables in the model predict *y,* using the prediction equation?
3. **The contingency table and the comparison using ratios**

The following table reports the distribution of a sample of Italian tourists (aged more than 18 years old) by the choice of holiday destination (Y) and Gender (X).

|  | **Male** | **Female** |
| --- | --- | --- |
| **Sea** | 180 | 140 |
| **Mountain** | 300 | 380 |

*Tab.2 A contingency table of the distribution of a sample of Italian tourists (aged more than 18 years old) by the choice of holiday destination (Y) and Gender (X).*

1. Are X and Y stochastically independent? Justify the answer.
2. Compare the proportions of those who went to the mountains with respect to gender. Considering the success as Y="Sea" and the failure as Y="Mountain", compute the odds ratio and comment on the result.

**3) The logistic regression and the interpretation of the parameters.**

The Table 3 reports the results on a logistic regression model, where:

* *Y* = 1 (pleasant/favorable service experience), 0 (unpleasant/unfavorable service experience)
* *X*1 = You were made to feel welcome
* *X*5 = Staff responded effectively to your inquiries
* *X*6 = Staff showed a sincere interest in solving your problem

|  |
| --- |
|  |
| **Variables in the equation** | **β** | **SE** | **Wald** | **df** | **Sig.** | **R** | **Exp (β)** |
| X1 Tourists were made to feel welcome | 1.1324 | 0.2256 | 25.1487 | 1 | 0.0001 | 0.2432 | 3.1034 |
| X5 Tourist information was clear and helpful | 0.6739 | 0.2534 | 7.0752 | 1 | 0.0089 | 0.1116 | 1.9621 |
| X6 Tourist amenities were satisfactory | 1.2187 | 0.2687 | 20.5638 | 1 | 0.0001 | 0.2104 | 3.3823 |
| Constant | -6.9543 | 0.9874 | 49.5539 | 1 | 0.0001 |  |  |

| **Observed** | **Unpleasant Experience** | **Pleasant Experience** | **% Correct** |
| --- | --- | --- | --- |
| Unpleasant experience | 83 | 20 | 80.6% |
| Pleasant experience | 18 | 210 | 92.1% |
| Overall |  |  | 88.5% |

Table 3: Results on a logistic regression model

1. Based on the logistic regression results, interpret the coefficients and the odds ratios for the variables X1, X5, and X6, assessing the statistical significance of the predictors.
2. Evaluate the accuracy and predictive power of the logistic regression model based on the provided results.

1. **The regression tree and the prediction the final value of Y for the corresponding leaf (cases: Qualitative Target Variable and Quantitative Target Variable)**

The Figure 1 reports the results on a regression tree, where:

* y- Global satisfaction (Quantitative Target Variable)
* X1- Work force of the respondent
* X2- Age of the respondent (in class)
* X3- Place of living
* ![Immagine che contiene testo, diagramma, Piano, Disegno tecnico

  Descrizione generata automaticamente]()X4-Occupation reliance in tourism

Immagine che contiene testo, schermata, diagramma, Rettangolo

Descrizione generata automaticamente

Figure 1: the output of a regression tree procedure.

1. Describe accurately the final leaves with respect to the target variable.
2. How do X1, X2, X3 and X4 contribute to predicting Y?