**EXAM**

**STATISTICAL METHODS FOR TERRITORIAL AND SOCIAL ENVIRONMENTS**

**21/06/2024**

1. **The type of multivariate relationships.**

A tourist city wants to analyze the factors that influence tourist satisfaction, the number of attractions visited, expenditure, and the length of stay. Data has been collected from a sample of 1000 tourists. Tourism managers want to understand how these variables are correlated to improve marketing strategies and resource management for tourism. Table 1 reports the correlations matrix between the variables.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Tourist Satisfaction** | **Number of Attractions Visited** | **Expenditure** | **Length of Stay** |
| **Tourist Satisfaction** | 1 |  |  |  |
| **Number of Attractions Visited** | 0.2 \*\* | 1 |  |  |
| **Expenditure** | 0.57 \*\* | 0.15 \*\* | 1 |  |
| **Length of Stay** | 0.35 \*\* | 0.2 \*\* | -0.27 \*\* | 1 |

*Tab. 1: Correlations matrix between the variables (\*\*=pvalue<0.0001)*

1. What are the main factors that influence tourist satisfaction?
2. Does the number of attractions visited by tourists significantly impact their satisfaction?
3. How is expenditure influenced by the length of stay and tourist satisfaction?
4. What factors mainly determine the length of stay of tourists?

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

The Table 2 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% |

*Tab. 2: 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 classification tree and the prediction of the final value of Y for the corresponding leaf**

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

* y- personal benefit from Tourism (Qualitative Target Variable)
* X1- Educational attainment (in group)
* X2- Occupational reliance on tourism
* X3- Place of living
* X4- Age (in group)

![A diagram of a graph

Description automatically generated with medium confidence]() ![A graph of a graph showing different stages of growth

Description automatically generated with medium confidence]()

*Figure 1: the output of a regression tree procedure.*

1. Describe accurately the final leaves with respect to the target variable. Moreover give the prediction of the final value of Y for the corresponding leaf
2. How do X1, X2, X3 and X4 contribute to predicting Y?

**4) The use of the qualitative covariates in the multiple linear regression.**

Table 4 reports the results of a multiple linear regression on 1000 tourism agencies to study the determinants respect to the annual revenue of the agency:

* Y= **Revenue**: The annual revenue generated by each tourism agency.
* X1=**Number of Tours Offered**: The annual number of different tours offered by the agency.
* X2= **Agency Location**: The location of the agency (e.g., Coastal, Urban, Rural)
* X3= **Type of Tours**: The main type of tours offered by the agency (e.g., Adventure, Cultural, Leisure).

| **Variable** | | | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | | | 30,100 | 470.023 | 64.049 | <0.0001 |
| Number of Tours Offered | | | 537.68 | 27.154 | 19.801 | <0.0001 |
| Location: Coastal | | | 9,259.37 | 419.284 | 22.084 | <0.0001 |
| Location: Urban | | | 4,771.52 | 384.232 | 12.418 | <0.0001 |
| Tours: Adventure | | | -9,918.04 | 409.288 | -24.232 | <0.0001 |
| Tours: Cultural | | | -3,230.34 | 382.327 | -8.449 | <0.0001 |
|  |
| **Component** | **df (Degrees of Freedom)** | **SS (Sum of Squares)** | |
| Regression | 5 | 6.79·(10^12) | |
| Residual | 994 | 6.49·(10^12) | |
| Total | 999 | 13.3·(10^12) | |

*Tab.3: 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?