

## Data Analysis

Final project consists of several statistical problems based on the statistical methods covered in this course.

- To receive maximum points, you need to have:
  - Accurate analysis
  - Appropriate tables/graphs and reports
  - Correct wording when reporting results
  - Appropriate interpretation of the results
  - Adequate formatting of the document
- Report should be submitted as a PDF document with 1-inch margins

Two datasets are used in the analysis:

### Dataset 1

Open **Hotel Room Design.sav** in SPSS and check all variables and examine data. Please test hypothesis and use 7-step approach.

### Part 1 - Multiple regression analysis - 50 points

- **Variables used in the analysis:**
  - SAT - Satisfaction - Average of Sat 1-7
  - Travel\_reason - What is the most common reason for your travel?
  - COLOR - Perceived quality of color - Average of COLOR 1, 2, and 3
  - FURN - Perceived quality of furniture - Average of FURN 1, 2, and 3
  - AEST - Perceived design quality - Average of AEST 1-7
  - LIGHT - Perceived quality of lighting - Average of LIGHT 1, 2, 3 and 4
  - LAYOUT - Perceived quality of layout - Average of LAYOUT 1, 2, 3, 4 and 5
- **Step 1:** Recognize research question and formulate hypotheses
  - Research Question 1 – Do most common reason for travel, and perceived quality of color, furniture, design, lighting, and layout predicts/explains customer satisfaction?
  - Research Question 2 for Multiple - Which variable among reason for travel, and perceived quality of color, furniture, design, lighting, and layout has the strongest effect on customer satisfaction?
  - Formulate Hypothesis
- **Step 2:** Determine how many variables you are comparing
  - Choose the Significance Level – common  $\alpha = 0.05$
  - Recognize independent and dependent variables
  - Report variable(s) and their types
  - Model Specification
- **Step 3: Data requirements and initial analysis**

- Dependent variable needs to be interval or ratio scaled (continuous)
- Independent variables need to be interval or ratio scaled (continuous) or converted to dummy variables
  - Convert categorical variables to dummy variables
- Run analysis
- Report sample size
- Do dependent and independent variables show variation
- Report (Multi)collinearity
- **Step 4:** Check Assumptions
  - Report assumptions
- **Step 5:** If the assumption are met proceed with model estimation
- **Step 6:** Interpretation and validation
  - Overall model fit
  - The effects of the independent variables separately (unstandardized coefficients)
  - The effects of the independent variables separately (standardized coefficients)
  - Validation - Split Sample - Randomly select 30 % of cases
    - Compare the model results
- **Step 7: Reporting results**

## Part 2 - Logistic regression analysis – 50 Points

Open **Hotel Room Design.sav** in SPSS and check all variables and examine data. Please test hypothesis and use 7 step approach.

- **Variables used in the analysis:**
  - PI\_bin - Purchase Intention Binary - Would you book this room?
  - Travel\_reason - What is the most common reason for your travel?
  - Gender
  - Age\_M - Customer Age Group
  - COLOR - Perceived quality of color - Average of COLOR 1, 2, and 3
  - FURN - Perceived quality of furniture - Average of FURN 1, 2, and 3
  - AEST - Perceived design quality - Average of AEST 1-7
  - LIGHT - Perceived quality of lighting - Average of LIGHT 1, 2, 3 and 4
  - LAYOUT - Perceived quality of layout - Average of LAYOUT 1, 2, 3, 4 and 5
- **Step 1:** Recognize research question and formulate hypotheses
  - Research Question 1 – Do most common reason for travel, gender, age group, and perceived quality of color, furniture, design, lighting, and layout predict which customers would book a room and which ones would not?
  - Research Question 2 for Multiple - Which variable among reason for travel, gender, age and perceived quality of color, furniture, design, lighting, and layout is the best predictor if customers would book a room and which ones would not?
  - Formulate Hypothesis

- **Step 2:** Determine how many variables you are comparing
  - Choose the Significance Level – common  $\alpha = 0.05$
  - Recognize independent and dependent variables
  - Report variable(s) and their types
  - Model Specification
- **Step 3: Data requirements**
  - Report sample size
  - Dependent variable needs to be dichotomous (binary categories)
  - Independent variables can be any type
  - Dependent and independent variables show variation
- **Step 4:** Check Assumptions
  - Generally, there are no assumptions
- **Step 5:** LR model estimation
- **Step 6:** Interpretation and validation
  - Overall model fit
  - Testing for significance of the coefficients – based on the Wald statistic
  - Interpreting the coefficients
    - Directionality of the relationship
    - Magnitude of the relationship of metric independent variables
  - Validation – Split Sample - Randomly select 30 % of cases
- **Step 7: Reporting results**

## Dataset 2

Open **Nightclub Price Fairness FINAL exam.sav** in SPSS and check all variables and examine data. Please test hypothesis and use 7 step approach where applicable.

This dataset includes information about customer's perception regarding pricing strategies for entrance fees that nightclubs use.

This study used 5x2 full-factorial experimental design.

Participants were randomly assigned to one of 10 scenarios and asked to imagine that they are customer in one of the scenarios:

Two variables were manipulated:

- Customer type based on how many times they have been to the nightclub in scenario (CustomerType). Participants were asked to imagine that they are one of the following two types:
  - Frequent customer
  - First time customer
- Pricing strategies (variable Pricing\_Strategy) with 5 conditions:
  - Day of the week – pricing strategy that calls for change in price of entrance fee depending on the day of the week (higher prices on weekend)

- Time of a day – pricing strategy that calls for change in price of entrance fee depending on the time when the customer shows up at the door (higher prices on late arrivals)
- Reservation in advance – pricing strategy that calls for change in price for customers that make reservations (higher prices for customers without reservation)
- VIP entrance – pricing strategy that calls for change in price for customers that do not wait in the line and those that wait in line (higher prices to go around the line)
- Flat pricing – pricing strategy that always includes the same price for all customers.

After learning about pricing strategy, participants were asked to answer a series of questions regarding their perception of the strategy.

They were asked to evaluate:

- Perceived price fairness – measured with 3 items (Fair1, Fair2 and Fair3)
- Word-of-mouth (would they recommend this nightclub to others) – measured with 3 items (WOM1, WOM2 and WOM3)
- Return Intention (would they return to this nightclub) – measured with 3 items (RI1, RI 2 and RI 3)
- Familiarity with the pricing strategy from scenario – measured with 2 items (FAM1 and FAM2)

After that participants were asked how often they go to nightclubs and a series of demographic questions (gender, age, ethnicity, income, employment status, and education)

### **Part 3 – Descriptives - 50 points**

- Create descriptive statistics for all demographic variables. Report means or frequencies depending on the type of variable and include graphs when applicable
  - Gender
  - Age
  - Ethnicity
  - Income
  - employment status
  - education
- Compute new variables:
  - Perceived\_Fairness - Average of Fair 1, 2 and 3
  - Word\_of\_mouth - Average of WOM 1, 2 and 3
  - Return\_Intention - Average of RI 1, 2 and 3
  - Familiarity - Average of FAM 1 and 2
- Create descriptive statistics for newly computed variables and report means and standard deviations.

#### **Part 4 – Hypothesis testing 1- 50 points**

- Use appropriate tests to test out following hypotheses:
  - H1: The average age is higher than 25 years.
  - H2: Frequent customers have higher word-of-mouth (use variable that you calculated for part 1) than first time customers.
  - H3: Female customers are more likely to be first-time customers compared to male customers that are more likely to be frequent customers
- Use 7-step approach and report results for each hypothesis.

#### **Part 5 – Hypothesis testing 2 – Analysis of Variance - 50 points**

- Conduct 5x2 full-factorial ANOVA
  - Test following hypotheses:
    - H4: First time customers have higher overall perceived price fairness compared to frequent customers.
    - H5: Customer type (Frequent customer vs. First time customer) moderates the relationship between type of pricing strategy and perceived price fairness (calculated in part1).
- Use 7-step approach and report results for each hypothesis.

#### **Part 6 – Hypothesis testing 3 – Regression - 50 points**

- Conduct regression analysis
- Build a model that would include following predictors of word-of-mouth
  - Age
  - Gender
  - Familiarity
  - Customer type
  - Perceived fairness
- Propose hypotheses for each potential relationship between independent and dependent variables
- Use 7-step approach and report results.