**Course Project: AJ DAVIS DEPARTMENT STORES**

|  |  |
| --- | --- |
|  |  |
| **Introduction** |  |

AJ DAVIS is a department store chain, which has many credit customers and wants to find out more information about these customers. A sample of 65 credit customers is selected with data collected on the following five variables.

1. Location (rural, urban, suburban)
2. Income (in $1,000's—be careful with the units)
3. Size (household size, meaning number of people living in the household)
4. Years (the number of years that the customer has lived in the current location)
5. Credit balance (the customers current credit card balance on the store's credit card, in $).

The data are available in Doc Sharing Course Project Data Set as an Excel file. You are to copy and paste the data set into a Minitab worksheet.

|  |  |
| --- | --- |
|  |  |
| **PROJECT PART A: Exploratory Data Analysis** |  |

* Open the file Copy of Math 533 Course Project Data – REV 7-2015.xlsx from the Course Project Data Set folder in Doc Sharing.
* For each of the five variables, process, organize, present, and summarize the data. Where appropriate, begin your analysis of each variable with Descriptive Statistics.
* Analyze each variable by itself using graphical and numerical techniques of summarization. Use Minitab as much as possible, explaining what the output tells you. You may wish to use some of the following graphs: stem-leaf diagram, frequency or relative frequency table, histogram, boxplot, dotplot, pie chart, bar graph. Caution: Not all of these are appropriate for each of these variables, nor are they all necessary. More is not necessarily better. In addition, be sure to find the appropriate measures of central tendency and measures of dispersion for the above data. Where appropriate use the five number summary (the Min, Q1, Median, Q3, Max). Once again, use Minitab as appropriate, and explain what the results mean.
* Analyze the connections or relationships between the variables. There are 10 possible pairings here (location and income, location and size, location and years, location and credit balance, income and size, income and years, income and balance, size and years, size and credit balance, years and Credit Balance). Use graphical as well as numerical summary measures. Explain what you see. Be sure to consider all 10 pairings. Some variables show clear relationships, while others do not.
* Prepare your report in Microsoft Word, ***integrating your graphs and tables with text explanations and interpretations.*** Be sure that you have graphical and numerical back up for your explanations and interpretations. Be selective in what you include in the report. I'm not looking for a 20-page report on every variable and every possible relationship (that's 15 things to do). Rather, what I want you do is to **highlight what you see for *three individual variables*** (no more than one graph for each, one or two measures of central tendency and variability (as appropriate), and two or three sentences of interpretation). For the 10 pairings, **identify and report only on *three of the pairings,*** again using graphical and numerical summary (as appropriate), with interpretations. ***Please note that at least one of your pairings must include location and at least one of your pairings must not include location****.*
* All DeVry University policies are in effect, including the plagiarism policy.
* Project Part A report is due by the end of Week 2.
* Project Part A is worth 100 total points. See grading rubric below.

**Submission: A Word document reporting the descriptive statistics and six graphs/charts described above, including all relevant graphs and numerical analysis along with interpretations**

**Format for report:**

1. Brief introduction
2. Table of Descriptive Statistics for any variables where these statistics are appropriate along with your interpretation
3. Discuss your first individual variable, using graphical, numerical summary, and interpretation
4. Discuss your second individual variable, using graphical, numerical summary, and interpretation
5. Discuss your third individual variable, using graphical, numerical summary, and interpretation
6. Discuss your first pairing of variables, using graphical, numerical summary, and interpretation
7. Discuss your second pairing of variables, using graphical, numerical summary, and interpretation
8. Discuss your third pairing of variables, using graphical, numerical summary, and interpretation
9. Conclusion

|  |  |
| --- | --- |
|  |  |
| **Project Part A: Grading Rubric** |  |

Getting the numerical analysis correct is certainly important but, since it’s done on the computer, this should be pretty straightforward. I’m really looking for an indication in your write-up that you understand what the analyses mean and how to interpret and use them.

| **Category** | **Points** | **%** | **Description** |
| --- | --- | --- | --- |
| Three Individual Variables  12 points each | 36 | 36 | graphical analysis, numerical analysis (when appropriate) and interpretation |
| Three Relationships  15 points each | 45 | 45 | graphical analysis, numerical analysis (when appropriate), and interpretation |
| Communication Skills | 19 | 19 | writing, grammar, clarity, logic, cohesiveness, adherence to the above format |
| **Total** | 100 | 100 | a quality paper will meet or exceed all of the above requirements. |

|  |  |
| --- | --- |
|  |  |
| **Project Part B: Hypothesis Testing and Confidence Intervals** |  |

Your manager has speculated the following.

a. The average (mean) annual income was less than $47,500.

b. The true population proportion of customers who live in an Urban area exceeds 40%.

c. The average (mean) number of years lived in the current home is less than 14 years.

d. The average (mean) credit balance for Suburban customers is more than $4,300.

**NOTE: be sure to select the correct test. Pay attention to what is being tested and the sample size.**

1. Using the sample data, perform the hypothesis test for each of the above situations in order to see if there is evidence to support your manager’s belief in each case A–D. In each case, use the Seven Elements of a Test of Hypothesis in Section 6.2 of your text book with α = .05, and explain your conclusion in simple terms. Also, be sure to compute the p-value and interpret.
2. Follow this up with computing 95% confidence intervals for each of the variables described in A–D, and again interpreting these intervals.
3. Write a report to your manager about the results, distilling down the results in a way that would be understandable to someone who does not know statistics. Clear explanations and interpretations are critical.
4. All DeVry University policies are in effect, including the plagiarism policy.
5. Project Part B report is due by the end of Week 6.
6. Project Part B is worth 100 total points. See the grading rubric below.

**Submission: A written report in Word and all of the relevant work done in the hypothesis testing (including Minitab) and the confidence intervals (Minitab) with your interpretation for each analysis.**

**Format for report:**

1. Summary report (about one paragraph on each of the speculations, A–D)
2. Include all of the steps in hypothesis testing (the format of the Seven Elements of a Test of Hypothesis, in Section 6.2 of your text book) for each speculation A–D, as well as the confidence intervals with written interpretations, including all Minitab output.

|  |  |
| --- | --- |
|  |  |
| **Project Part B: Grading Rubric** |  |

| **Category** | **Points** | **%** | **Description** |
| --- | --- | --- | --- |
| Addressing each hypothesis—20 points each | 80 | 80 | hypothesis test, interpretation, confidence interval, and interpretation |
| Summary report | 20 | 20 | one paragraph on each of the speculations |
| **Total** | 100 | 100 | A quality paper will meet or exceed all of the above requirements. |

|  |  |
| --- | --- |
|  |  |
| **Project Part C: Regression and Correlation Analysis** |  |

Using Minitab, perform regression and correlation analysis for the data on income, the independent variable X, vs. credit balance, the dependent variable Y, by answering the following:

1.        Generate a scatterplot for income ($1,000) versus credit balance ($), including the graph of the best fit line. Interpret.

2.        Determine the equation of the best fit line, which describes the relationship between income and credit balance.

3.        Determine the coefficient of correlation. Interpret.

4.        Determine the coefficient of determination. Interpret.

5.        Test the utility of this regression model (use a two tail test with α =.05). Interpret your results, including the p-value.

6.        Based on your findings in 1–5, what is your opinion about using income to predict credit balance? Explain.

7.        Compute the 95% confidence interval for beta-1 (the population slope). Interpret this interval.

8.        Using an interval, estimate the average credit balance for customers who have an income of $45K. Interpret this interval.

9.        Using an interval, predict the credit balance for a customer who has an income of $45K. Interpret this interval.

10.       What can we say about the credit balance for a customer who has an income of $100,000? Explain your answer.

In an attempt to improve the model, we decide to run a multiple regression model predicting credit balance based on income, years, and size.

11.          Using Minitab, run the multiple regression analysis using the variables income, years, and size to predict credit balance. State the equation for this multiple regression model.

12.          Perform the global test for Utility (F-Test). Explain your conclusion.

13.          Perform the t-test on each independent variable. Explain your conclusions and clearly state how you should proceed. In particular, state which independent variables we should keep and which should be discarded.

14.          Is this multiple regression model better than the linear model that we generated in parts 1–10? Explain.

All DeVry University policies are in effect, including the plagiarism policy.

15.  Project Part C report is due by the end of Week 7.

16.  Project Part C is worth 100 total points. See the grading rubric below.

**Summarize your results from 1–14 in a report that is 3 pages or less in length and explains and interprets the results in ways that are understandable to someone who does not know statistics.**

**Submission: The summary report + all of the work done in 1–14 (Minitab Output + interpretations) as an appendix**

**Format:**

1. Summary Report
2. Points 1–14 addressed with appropriate output, graphs, and interpretations. **Be sure to number each point 1–14.**

|  |  |
| --- | --- |
|  |  |
| **Project Part C: Grading Rubric** |  |

| **Category** | **Points** | **%** | **Description** |
| --- | --- | --- | --- |
| Questions 1–12 and 14  5 points each | 65 | 65 | addressed with appropriate output, graphs, and interpretations |
| Question 13 | 15 | 15 | addressed with appropriate output, graphs, and interpretations |
| Summary | 20 | 20 | writing, grammar, clarity, logic, and cohesiveness |
| **Total** | 100 | 100 | A quality paper will meet or exceed all of the above requirements. |

|  |  |
| --- | --- |
|  |  |
|  |  |