**MBA 693**

**Problem 1 (60 points)**

This concerns a bank's efforts to reduce the rate of loan defaults. As a member of the financial services analytics team at BSA Bank, you are assigned the task to assist a loan officer who needs to determine customer credit risk level based on customer demographics and financial stability.

Information on 100 past customers is contained in *bankloan\_A.jmp* (located in the Week 10 Module on Canvas).

The data set includes the following variables:

1. **Risk \_Score:** Credit risk level score (the higher the score, the riskier)
2. **Age:** customer age in years
3. **College:** an indicator of whether the customer has a college education (1 = yes, 0 = no)
4. **Employment:** years that the customer has been with his/her current employer
5. **Address:** years that the customer has lived at his/her current address
6. **Debt \_to \_Income:** debt to income ratio (x100)
7. **Other \_Debt:** other debt (in $1,000)

Using JMP Pro and *bankloan\_A.jmp,* generate **multiple** linear regression/correlation analysis to assess customer age, customer college education, years that the customer has been with his/her current employer, years that the customer has lived at his/her current address, customer debt-to-income ratio, and other debt as predictors of customer credit risk level score. Summarize your findings of the multiple linear regression/correlation analysis. What would you say is the **best** **model** for predicting customer credit risk level score? Be sure to back up your summary with evidence; include JMP Pro output, as necessary.

Document your complete, thorough analysis on the *Exam Worksheet:* **Problem 1: Bank Loan.** Be sure to include the JMP Pro output in the Appendix.

**Problem 2 (40 points)**

This concerns a health insurance company’s effort to set the price of patient yearly premiums. Health insurance companies must collect more in yearly premiums than it spends on medical care for its beneficiaries so that they can stay afloat. In order to set the price of patient yearly premiums, which can be high or low, depending on expected treatment costs, health insurance companies need to determine patient medical care expenses based on patient characteristics.

You are assigned the task to assist HCA Insurance in their effort to set the price of yearly premiums for its insured population.

Data on 86 patients serviced by HCA Insurance is obtained. The data include the following:

1. **expenses**: patient medical care expenses ($)
2. **age**: age of patient, in years
3. **gender**: gender of patient (0 = male, 1 = female)
4. **bmi**: patient body mass index
5. **children**: patient number of dependents

Refer to the JMP Pro output given in the Appendix*.*

Answer the questions on the *Exam Worksheet:* **Problem 2: HCA Insurance**

**a**. How strong is the correlation between patient medical care expensesand patient body mass index? Be sure to defend your answer.

**b**. How much of the variation in patient medical care expensesis explained for by the variation in patient body mass index?

**c**. Is patient body mass index a *significant* predictor of patient medical care expenses? Be sure to defend your answer.

**d**. Any concerns about the validity of a regression model that uses patient body mass index as a predictor of patient medical care expenses? Be sure to defend your answer.

**e**. Assess the *usefulness* and *significance* of patient age as a predictor of patient medical care expenses*.* Be sure to defend your answer.

**f**. Which of the two predictors, patient body mass index or patient age, is the single **best** predictor of patient medical care expenses? Defend your answer.

**g**. A patient is 42 years of age and has a body mass index of 30.2. Using the single best predictor identified in part (f), predict medical care expensesfor this patient.

**Appendix**

**JMP Pro Output: HCA Insurance**

**Distributions**

**expenses**



**Quantiles**

|  |  |  |
| --- | --- | --- |
| 100.0% | maximum | 54200.99 |
| 75.0% | quartile | 44295.665 |
| 50.0% | median | 37149.53 |
| 25.0% | quartile | 24308.01 |
| 0.0% | minimum | 16577.78 |

**Summary Statistics**

|  |  |
| --- | --- |
| Mean | 35163.655 |
| Std Dev | 10957.57 |
| N | 86 |

**bmi**



**Quantiles**

|  |  |  |
| --- | --- | --- |
| 100.0% | maximum | 44.5 |
| 75.0% | quartile | 38.1 |
| 50.0% | median | 33.2 |
| 25.0% | quartile | 27.875 |
| 0.0% | minimum | 19.8 |

**Summary Statistics**

|  |  |
| --- | --- |
| Mean | 32.94186 |
| Std Dev | 6.2540427 |
| N | 86 |

**age**



**Quantiles**

|  |  |  |
| --- | --- | --- |
| 100.0% | maximum | 64 |
| 75.0% | quartile | 51 |
| 50.0% | median | 42 |
| 25.0% | quartile | 27 |
| 0.0% | minimum | 18 |

**Summary Statistics**

|  |  |
| --- | --- |
| Mean | 39.976744 |
| Std Dev | 13.674344 |
| N | 86 |

**Multivariate Correlations**

|  | **expenses** | **bmi** | **age** |
| --- | --- | --- | --- |
| expenses | 1.0000 | 0.9184 | 0.2693 |
| bmi | 0.9184 | 1.0000 | 0.1102 |
| age | 0.2693 | 0.1102 | 1.0000 |

**Scatterplot Matrix**



**Model 1:**

**Bivariate Fit of expenses by bmi**



**Linear Fit**

expenses = -17845.69 + 1609.1789\*bmi

**Summary of Fit**

|  |  |
| --- | --- |
| RSquare | 0.843533 |
| RSquare Adj | 0.84167 |
| Root Mean Square Error | 4360.096 |
| Mean of Response | 35163.65 |
| Observations (or Sum Wgts) | 86 |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
| --- | --- | --- | --- | --- |
| Model | 1 | 8608931987 | 8.6089e+9 | 452.8529 |
| Error | 84 | 1596876943 | 19010440 | **Prob > F** |
| C. Total | 85 | 1.0206e+10 |  | <.0001\* |

**Parameter Estimates**

| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| --- | --- | --- | --- | --- |
| Intercept | -17845.69 | 2534.983 | -7.04 | <.0001\* |
| bmi | 1609.1789 | 75.6181 | 21.28 | <.0001\* |

**Diagnostics Plots**

**Residual by Predicted Plot**



**Actual by Predicted Plot**



**Residual by Row Plot**



**Residual by X Plot**



**Residual Normal Quantile Plot**



**Durbin-Watson**

| **Durbin-Watson** | **Number of Obs.** | **AutoCorrelation** |
| --- | --- | --- |
| 2.0613047 | 86 | -0.0375 |

**Model 2:**

**Bivariate Fit of expenses by age**



**Linear Fit**

expenses = 26538.401 + 215.75678\*age

**Summary of Fit**

|  |  |
| --- | --- |
| RSquare | 0.072496 |
| RSquare Adj | 0.061454 |
| Root Mean Square Error | 10615.54 |
| Mean of Response | 35163.65 |
| Observations (or Sum Wgts) | 86 |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
| --- | --- | --- | --- | --- |
| Model | 1 | 739879207 | 739879207 | 6.5656 |
| Error | 84 | 9465929723 | 112689640 | **Prob > F** |
| C. Total | 85 | 1.0206e+10 |  | 0.0122\* |

**Parameter Estimates**

| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob>|t|** |
| --- | --- | --- | --- | --- |
| Intercept | 26538.401 | 3555.462 | 7.46 | <.0001\* |
| age | 215.75678 | 84.2027 | 2.56 | 0.0122\* |

**Diagnostics Plots**

**Residual by Predicted Plot**



**Actual by Predicted Plot**



**Residual by Row Plot**



**Residual by X Plot**



**Residual Normal Quantile Plot**

