**MSCR 500/533 Final Exam Report on Oxygen Consumption**

**INTRODUCTION**

In an exercise physiology study, the relationship between oxygen consumption was tested against several predictors (age, sex, pulse rate, perceived physical fitness/performance, and time required to run 1.5 miles). 31 subjects were sampled.

Here are the sample variable descriptons:

- Oxygen consumption (% of available oxygen consumed, as measured during 1.5 mile treadmill run. The more fit a person is, the greater their oxygen consumption will be during physical activity)

- Sex (M, F)

- Age (years)

- Weight (kg)

- Run time (time, in minutes, taken to run 1.5 miles on treadmill)

- Maximum pulse (highest pulse rate during the run)

- Performance (a subjective measure of the subject’s fitness, assigned by an interviewer; larger values mean the subject was assessed as having a higher fitness level)

The data was analyzed with the intent to create a predictive regression model assessing oxygen consumption.

**DESCRIPTION**

Here is a table showing the first 10 observations made, as an example of the data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Obs** | **Name** | **Age** | **Sex** | **Weight** | **Oxygen\_Consumption** | **RunTime** | **Performance** | **Maximum\_Pulse** |
| **1** | Donna | 42 | F | 68.15 | 59.57 | 8.17 | 90 | 172 |
| **2** | Gracie | 38 | F | 81.87 | 60.06 | 8.63 | 94 | 186 |
| **3** | Luanne | 43 | F | 85.84 | 54.3 | 8.65 | 83 | 168 |
| **4** | Mimi | 50 | F | 70.87 | 54.63 | 8.92 | 67 | 155 |
| **5** | Chris | 49 | M | 81.42 | 49.16 | 8.95 | 72 | 185 |
| **6** | Allen | 38 | M | 89.02 | 49.87 | 9.22 | 92 | 180 |
| **7** | Nancy | 49 | F | 76.32 | 48.67 | 9.4 | 64 | 188 |
| **8** | Patty | 52 | F | 76.32 | 45.44 | 9.63 | 56 | 166 |
| **9** | Suzanne | 57 | F | 59.08 | 50.55 | 9.93 | 43 | 155 |
| **10** | Teresa | 51 | F | 77.91 | 46.67 | 10 | 54 | 168 |

To describe the data, here are descriptive tables of the continuous and categorical variables.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **Age** | 31 | 47.67742 | 5.26236 | 1478 | 38.00000 | 57.00000 |
| **Maximum\_Pulse** | 31 | 173.77419 | 9.16410 | 5387 | 155.00000 | 192.00000 |
| **Performance** | 31 | 56.64516 | 18.32584 | 1756 | 20.00000 | 94.00000 |
| **RunTime** | 31 | 10.58613 | 1.38741 | 328.17000 | 8.17000 | 14.03000 |
| **Weight** | 31 | 77.44452 | 8.32857 | 2401 | 59.08000 | 91.63000 |
| **Oxygen\_Consumption** | 31 | 47.37581 | 5.32777 | 1469 | 37.39000 | 60.06000 |

|  |  |  |
| --- | --- | --- |
| **Sex** | **Frequency** | **Percent** |
| **F** | 16 | 51.61 |
| **M** | 15 | 48.39 |

The outcome variable, oxygen consumption, was assessed in further detail for linearity and normality.





Based on the histogram and probability plot, oxygen consumption was both normally distributed and showed a linear pattern, allowing for its use in a linear model.

**FINAL MODEL**

The final regression model chosen:

This final model was chosen through clinical parameters. Oxygen consumption is based on an overall cardiopulmonary fitness. The most powerful and predictive measure should be the physical activity level (e.g. stress/treadmill test). The other factors that have a clinical impact on oxygen consumption are sex (difference at baseline physiological use of oxygen/extraction), age (differences in metabolism), and weight (overall metabolic burden). Perceived performance was excluded due to its subjective assessment and its strong correlation with RunTime (as seen below, with model diagnostics and development).

The final model above, had an f-statistic of 32.16. The p-value for the overall model was p < 0.0001. The adjusted R-square for the model was 75.70%. With this model, every 1 minute increase in 1.5 mile runtime leads to a 2.96% decrease in oxygen consumption.

The parameter estimates for the individual variables are described in the table below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **Parameter Estimate** | **Standard Error** | **t-statistic** | **p-value** | **95% Confidence Limits** | |
| **B0** | Intercept | 87.55667 | 9.18045 | 9.54 | <.0001 | 68.68599 | 106.42734 |
| **B1** | Sex | -1.41906 | 1.31662 | -1.08 | 0.2910 | -4.12541 | 1.28729 |
| **B2** | Age | -0.15768 | 0.09988 | -1.58 | 0.1265 | -0.36300 | 0.04763 |
| **B3** | Weight | -0.00783 | 0.07616 | -0.10 | 0.9189 | -0.16438 | 0.14872 |
| **B4** | RunTime | -2.96328 | 0.39868 | -7.43 | <.0001 | -3.78279 | -2.14378 |

**DIAGNOSTICS AND DEVELOPMENT**

In developing the model, the first step was checking for correlation both visually and through statistical testing. Visually, scatter plots were generated for each relationship.



The strongest visual relationship was with RunTime. This is seen as well in the correlation table below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Age** | **Oxygen\_Consumption** | **Weight** | **NumSex** | **Maximum\_Pulse** | **Performance** | **RunTime** |
| **Age** | 1.0000 | -0.3116 | -0.2405 | 0.0603 | -0.4149 | -0.7126 | 0.1952 |
| **Oxygen\_Consumption** | -0.3116 | 1.0000 | -0.1629 | -0.4845 | -0.2368 | 0.7789 | -0.8622 |
| **Weight** | -0.2405 | -0.1629 | 1.0000 | 0.5718 | 0.2494 | 0.0897 | 0.1435 |
| **NumSex** | 0.0603 | -0.4845 | 0.5718 | 1.0000 | 0.2176 | -0.2531 | 0.4312 |
| **Maximum\_Pulse** | -0.4149 | -0.2368 | 0.2494 | 0.2176 | 1.0000 | 0.0900 | 0.2261 |
| **Performance** | -0.7126 | 0.7789 | 0.0897 | -0.2531 | 0.0900 | 1.0000 | -0.8205 |
| **RunTime** | 0.1952 | -0.8622 | 0.1435 | 0.4312 | 0.2261 | -0.8205 | 1.0000 |

Sex was included as a variable due to 1) norms in biostatistical research, and 2) the relationship based on t-test with oxygen consumption, and 3) underlying physiological differences. Similar tests were done for the other variables. The t-test for sex by oxygen consumption showed a t-statistic of 2.99, with a p-value = 0.0056.



Runtime was thought to be the most important predictive factor, and thus assessed and visualized as well.



This figure shows this linearity pattern, and emphasizes its utility in our final model.