AN ANALYSIS OF IMPORTANT FACTORS ON DETERMINING EXERCISE TOLERANCE

A PROJECT REPORT
Presented to the Department of Mathematics and Statistics
California State University, Long Beach

In Partial Fulfillment of the Requirements for the Degree
Master of Science in Applied Statistics

Faculty Reviewer:

Tianni Zhou, Ph.D.

By:
Kalvin Ogbuefi
M.S., 2018, California State University, Long Beach
Russell Valenzuela
M.S., 2018, California State University, Long Beach

May 2017

ABSTRACT

The purpose of the experiment was to understand which factors were significant in influencing exercise tolerance. Three main factors of interest included gender, body fat, and smoking history. Analysis was done using the statistical software, SAS, through a generalized linear model procedure. The results illustrate in detail the significance of influencing factors gender, body fat level, and smoking history along with the interaction between body fat level and smoking history at the level of α= 0.05. The analysis concludes by summarizing why each factor played a role in being either significant or not to influencing exercise tolerance.

INTRODUCTION

Many factors influence an individual's ability to tolerate or withstand exercise. It is wildly accepted a person's overall health is a defining factor in determining their endurance level in performing physical tasks. In general, those with less body fat tend to live healthier lives and are less at risk for cardiovascular diseases. Thus, those individuals should also be able to perform at a higher level when performing a specific exercise. Likewise, previous studies have linked smoking history to respiratory problems. As for gender, due to the physiological makeup of both males and females, one may argue that one's gender should have an advantage or disadvantage in performing certain physical tasks. All other factors included with one's overall health aid in determining if this is significant or not.

The goal of the experiment is to understand how gender (A), body fat level (B), and smoking history (C) influence exercise tolerance. The data were collected with three subjects from each gender-body fat-smoking history group riding a bicycle apparatus. The number of minutes each subject could tolerate or perform the bicycle exercise were recorded with three replications. Each of the factors have two levels with gender being either male or female, body fat level being low or high, and smoking history being light or heavy. In addition to determining which of the three factors significantly impact exercise tolerance, the experiment also aims to quantify the effects the factors may have.

METHODS

When wanting to study the effects of two or more factors, factorial designs are generally used. Each factor in the experiment will have a number of levels. For example, Factor *A* will have *a* levels, Factor *B* will have *b* levels, and so on. If all the factors of interest have 2 levels, the 2^k factorial design can be used. The total number of treatment combinations used in the experiment is then 2^k, where *k* is the number of factors. In the exercise tolerance experiment, there are three factors as previously stated. As a result, all proceeding equations and figures are for a 2^3 factorial design with eight treatments. The design has the following model:

$$y_{ijmn} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijm} + \epsilon_{ijmn}$$

$$i = 1, 2$$

$$j = 1, 2$$

$$m = 1, 2$$

$$n = 1, 2, 3$$

$$\mu\text{: } overall\ mean$$

$$\tau_i\text{: } Effect\ from\ ith\ level\ of\ Factor\ A$$

$$\beta_j\text{: } Effect\ from\ jth\ level\ of\ Factor\ B$$

$$\gamma_m\text{: } Effect\ from\ kth\ level\ of\ Factor\ C$$

$$(\tau\beta)_{ij}\text{: } Interaction\ Effect\ from\ ith\ level\ of\ Factor\ A\ and\ jth\ level\ of\ Factor\ B$$

$$(\tau\gamma)_{im}\text{: } Interaction\ Effect\ from\ ith\ level\ of\ Factor\ A\ and\ kth\ level\ of\ Factor\ C$$

$$(\beta\gamma)_{jm}\text{: } Intereaction\ Effect\ from\ jth\ level\ of\ Factor\ B\ and\ kth\ Level\ of\ Factor\ C$$

$$(\tau\beta\gamma)_{ijm}\text{: } Intereaction\ Effect\ from\ ith\ level\ of\ Factor\ A, jth\ level\ of\ Factor\ B, kth\ level\ of\ Factor\ C$$

$$\epsilon_{ijmn}\text{: } Random\ Error,$$

under the following constraints:

$$\tau_1 + \tau_2 = 0, \qquad \beta_1 + \beta_2 = 0, \qquad \gamma_1 + \gamma_2 = 0,$$

$$\sum_{i=1}^{2} (\tau\beta)_{ij} = \sum_{j=1}^{2} (\tau\beta)_{ij} = 0, \qquad \sum_{i=1}^{2} (\tau\gamma)_{im} = \sum_{m=1}^{2} (\tau\gamma)_{im} = 0, \qquad \sum_{j=1}^{2} (\beta\gamma)_{jm} = \sum_{m=1}^{2} (\beta\gamma)_{jm} = 0,$$

$$\sum_{i=1}^{2} (\tau\beta\gamma)_{ijm} = \sum_{j=1}^{2} (\tau\beta\gamma)_{ijm} = \sum_{m=1}^{2} (\tau\beta\gamma)_{ijm} = 0$$

and the assumption:

$$\epsilon_{ijmn} \sim N(0, \sigma^2)$$

Determining the necessary parameter estimates or effects and sum of squares of can be found by finding the contrasts for the main effects A, B, and C as well as the two and three way interactions between the three. Table 1 displays the algebraic signs used for computing contrasts. The two left columns represent the eight treatment combinations in Yates Order and their totals from the dataset. The presence of a letter indicates the 2nd/high level of the corresponding factor.

Treatment Combination	Total	I	A	B	C	AB	AC	BC	ABC
(1)	77.9	+	-	-	-	+	+	+	-
a	59.5	+	+	-	-	-	-	+	+
b	42.2	+	-	+	-	-	+	-	+
ab	36.2	+	+	+	-	+	-	-	-
c	59.6	+	-	-	+	+	-	-	+
ac	36.4	+	+	-	+	-	+	-	-
bc	48.1	+	-	+	+	-	-	+	-
abc	30.6	+	+	+	+	+	+	+	+

Table 1: Sign Table for Finding Effect Contrasts in 2^3 Design

From Table 1, the computational form for the contrast of A is given as:

$$Contrast\ A = -(1) +\ a - b + ab - c + ac - bc + abc$$

For B:

$$Contrast\ B = -(1) - a + b + ab - c - ac + bc + abc$$

and so on.

Given the contrasts, the effects and sum of squares can be computed as follows:

$$Effect = \frac{Contrast}{2^{k-1}n} = \frac{Constrast}{12}$$

$$SS = \frac{Contrast}{2^k n} = \frac{Constrast}{24}$$

Additionally, the following computes total sum of squares and error sum of squares:

$$SS_{Total} = \sum (individual\ observations)^2 - \frac{(grand\ total)^2}{24}$$

$$SS_{Error} =\ SS_{Total} - \sum SS_{Effect}$$

The degrees of freedom associated for each of the main and interaction effects is 1. Since there is a total of 23 degrees of freedom (24 observations – 1), the degrees of freedom associated for error is then 23-7=16. Knowledge of degrees of freedom allows for the mean squares to be calculated by dividing the sum of squares by the associated degrees of freedom.

To test whether an effect is significant and there is a difference between the two levels of a factor (H_0: Main/Interaction Effect = 0), an F-test is conducted with the following test statistic:

$$F_0 = \frac{MS_{Effect}}{MS_{Error}}$$

If F_0 for an effect is greater than $F_{(0.95,1,23)}$ =4.28, the effect is considered significant. Another method for determining an effect's significance is to create a confidence interval for the effect. Since the estimated variance of an effect will follow a t-distribution with the same degrees of freedom associated to error, 16, the lower and upper limits of a 95% confidence interval is found by:

$$Effect \pm t_{0.025,16} se(effect) = Effect \pm (2.120)\sqrt{\frac{MS_{Error}}{6}}$$

If the interval contains zero, the effect is considered insignificant.

RESULTS

FACTOR LEVEL MEANS

An initial exploratory analysis of the dataset produces the level means presented in Tables 2 through 6 and an overall mean of 16.270 minutes. From Table 2, males have a higher tolerance than females, individuals with low body fat have a higher tolerance than those with high body fat, and individuals with light smoking history have a higher tolerance than those with heavy smoking history. Table 3 indicates males with low body fat have the highest tolerance while females with high body fat have the lowest. Similarly, Table 4 shows males with light smoking history produce a high tolerance and females with high smoking history produce a low tolerance. The level means from Table 5 and tolerance are high for individuals with low body fat and light smoking history. Interestingly for individuals with high body fat, the tolerance is low regardless of the level of smoking history. Similar outcomes can be drawn from Table 6.

Factor	Level	Mean (mins)
A (Gender)	Male	18.983
	Female	13.558
B (Body Fat Level)	Low	19.450
	High	13.092
C (Smoking History)	Light	17.983
	Heavy	14.558

Table 2: Level Means for Factors A, B, and C

Gender Level	Body Fat Level	Mean (mins)
Male	Low	22.912
	High	15.050
Female	Low	15.983
	High	11.133

Table 3: Level Means for Two-Way Interaction Between A and B

Gender Level	Smoking History Level	Mean (mins)
Male	Light	20.167
	Heavy	17.950
Female	Light	15.950
	Heavy	11.167

Table 4: Level Means for Two-Way Interaction Between A and C

Body Fat Level	Smoking History Level	Mean (mins)
Low	Light	22.900
	Heavy	16.000
High	Light	13.067
	Heavy	13.117

Table 5:Level Means for Two-Way Interaction Between B and C

Gender Level	Body Fat Level	Smoking History Level	Mean (mins)
Male	Low	Light	25.967
		Heavy	19.867
	High	Light	14.067
		Heavy	16.033
Female	Low	Light	19.833
		Heavy	12.133
	High	Light	12.067
		Heavy	10.200

Table 6: Level Means for Three-Way Interaction Between A, B, and C

ANALYSIS OF VARIANCE (ANOVA)

While there are differences between the level means presented in Tables 2 through 6, the quantified main and interaction effects are shown in Table 7. As shown, gender, body fat level, and smoking history do have strong effects on exercise tolerance minutes. The effect estimates are for the 2^{nd}/high level of a factor (i=2, j=2, and m=2) which corresponds to female, high body fat level, and the high smoking history. On average, females are estimated to tolerate exercise stress by 5.425 fewer minutes. Likewise, individuals with high body fat have an estimated effect of 6.358 minutes and those with high smoking history have an estimate of 3.425 minutes. The ANOVA Table generated through SAS displayed in Table 8 shows the three main effects are significant as is the BC interaction (between body fat level and smoking history) at the α=0.05 significance level. Simultaneously, the AB, AC, ABC interactions are insignificant. An important outcome comes in the p-value for body fat is the lowest indicating it may be the strongest factor. The other p-values for the significant terms are similarly small and display an importance.

Source of Variation	Contrast	Effect Estimate	Sum of Squares
A (Gender)	-65.1	-5.425	176.58
B (Body Fat Level)	-76.3	-6.358	242.57
C (Smoking History)	-41.1	-3.425	70.38
AB	18.1	1.508	13.65
AC	-16.3	-1.358	11.07
BC	41.7	3.475	72.45
ABC	-6.7	-0.558	1.87

Table 7: Computed Contrasts, Effect Estimates, and Sum of Squares

Source	Degrees of Freedom	Sum Squares	Mean Square	F Value	p-value
A (Gender)	1	176.58	176.58	18.92	0.0005
B (Body Fat Level)	1	242.57	242.57	25.98	0.0001
C (Smoking History)	1	70.38	70.38	7.54	0.0144
AB	1	13.65	13.65	1.46	0.2441
AC	1	11.07	11.07	1.19	0.2923
BC	1	72.45	72.45	7.76	0.0132
ABC	1	1.87	1.87	0.20	0.6604
Error	16	149.37	9.34		
Total	23				

Table 8: ANOVA Table for Full Model Showing AB, AC, ABC interactions insignificant at the α=0.05 significance level

Similarly, the construction of 95% confidence intervals for each of the effects confirm the results drawn from the ANOVA Table. Since the confidence intervals for A, B, C, and BC do not contain zero, they are significant. The insignificant effects, AB, AC, and ABC contain zero.

Source of Variation	Lower Limit	Upper Limit
A (Gender)	-8.07	-2.78
B (Body Fat Level)	-9.00	-3.71
C (Smoking History)	-6.07	-0.78
AB	-1.14	5.15
AC	-4.00	1.29
BC	0.83	6.12
ABC	-3.20	2.09

Table 9: 95% Confidence Intervals for Effects showing same results from Table 8

TESTING FOR MODEL ASSUMPTIONS

An analysis of the model residuals hint at a minor departure from the assumptions of normally distributed residuals with a constant variance, σ^2. Figure 1 displays the residuals plotted against the model predicted values and the probability plot for the residuals. Because the plot of residuals against the fitted values do not necessarily display the same level of dispersion across, some assumption violation exists. Similarly, the probability plot does not completely follow a straight 45-degree line and as a result, may not be normally distributed as assumed by the model.

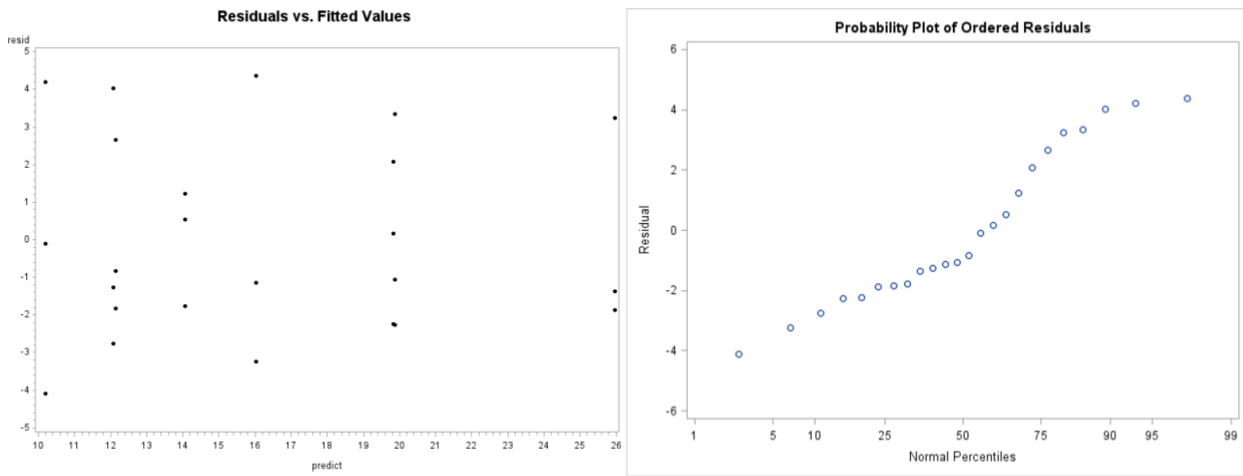

Figure 1: Residual Checking of ANOVA Model Assumptions showing some departure from constant variance and normality (L: Residuals vs. Fitted Values, R: Probability Plot of Ordered Residuals)

Despite minor model assumption violations, the results from the ANOVA Table can still be assumed to be adequate. The purpose of the experiment was to determine the influence of the three factors in question. The p-values associated to the main effects of these factors were low enough to indicate a strong significance with no reason to believe otherwise with a modified model.

Conclusion

Since the ANOVA Table indicates the main effects of factors A, B, and C as well as the interaction between B and C are significant, gender, body fat level, and smoking history all play an important part in impacting exercise tolerance. The effect of gender can be attributed to the different physiological makeup between males and females with males being able to tolerate exercise stress by nearly five and half minutes better than their female counterparts. As expected the individuals with low body fat and light smoking history lasted longer on the bicycle and were significantly able to endure exercise for longer period. The cardiovascular and respiratory impacts appear to come into play. However, the experiment makes no specific association and only assumes the causation of body fat and smoking history on a person's general health.

Additionally, the experiment design is somewhat limited in scope. A better analysis on body fat level can be studied in a further experiment. In the dataset explored, body fat only comprised of two levels, low and high. Since body fat level showed significance, using different and more levels for the factor is appropriate. Getting a better idea of how body fat level affects exercise tolerance since it displayed the highest importance. Regardless, the experiment still illustrated the impact gender, body fat level, and smoking history all have in an individual's ability to perform physical activity.