Journal of Statistics Education

ISSN: (Print) 1069-1898 (Online) Journal homepage: https://www.tandfonline.com/loi/ujse20

Analyzing Exercise Training Effect and Its Impact on Cardiorespiratory and Cardiovascular Fitness

Paul J. Laumakis & Kevin McCormack

Analyzing Exercise Training Effect and Its Impact on Cardiorespiratory and Cardiovascular Fitness

Paul J. Laumakis
Rowan University

Kevin McCormack
U.S. Postal Service

Journal of Statistics Education Volume 22, Number 2 (2014),
www.amstat.org/publications/jse/v22n2/laumakis.pdf

Key Words: Post exercise heart rate recovery; Multiple regression; Interaction model; Matched-sample hypothesis testing

Abstract

This paper provides a statistical investigation of the impact of heart rate levels on training effect for a specific exercise regimen, including an analysis of post-exercise heart rate recovery. Results indicate optimum target values for both average and maximum heart rate during exercise in order to improve both cardiorespiratory and cardiovascular fitness levels. The statistical methods used in the analysis are typically covered in college level Statistics I & II courses, and various classroom implementation strategies are presented.

1. Introduction and Story

Most teachers of statistics would agree that the use of real data in the classroom not only allows for the various descriptive and inferential statistical techniques being taught to be illustrated, but also serves to motivate students. On many occasions during my teaching career, I have personally witnessed that students will naturally tend to be more engaged with statistical analysis when the data that they are analyzing is drawn from real life experiences. Indeed, the American Statistical Association's (ASA's) Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report (2010) recommends the use of real data in the statistics classroom, while other authors (e.g. Hand, Daly, Lunn, McConway, and Ostrowski 1996) conclude that the use of real data demonstrates to students that statistics is about solving real life problems. Additionally, beyond recommending the use of real data in the statistics classroom, the ASA's GAISE College Report also endorses a fostering of active learning in the statistics classroom where the activity chosen begins and ends with an overview of what is being done and why, along with the use of technology for developing conceptual understanding and analyzing data.

The purpose of this article is to show how to utilize data in the statistics classroom that resulted from a personal quest to achieve a healthier lifestyle by Kevin, one of this article's authors. The data that is analyzed in the subsequent sections of this paper is real life data, collected by Kevin on a series of four-mile runs over a period of seven months.

After losing a couple friends to smoking-related cancer at relatively young ages, Kevin decided to make some positive changes in his own life at the age of 47. Not only did he commit to give up his casual cigarette smoking, but he also decided to begin a running program. Up to this point in his life, Kevin had never really exercised regularly and he was approximately 55 pounds overweight. He began his running program slowly, intermittently jogging slowly and walking around his neighborhood. After a period of time, he was able to run for relatively short distances and he eventually built up his stamina enough to run a four-mile loop near his home. In order to track his progress, Kevin decided to purchase a Global Positioning System (GPS) watch that would monitor his training data, including his heart rate, running pace, and calories burned, among other variables. After collecting training data for nineteen different runs, Kevin wants to analyze the data in order to see how he is progressing in his exercise program. Specifically, Kevin is interested in answering the following questions related to his health based on his running program:

1. How effective are his individual runs at improving his overall cardiorespiratory fitness?
2. To what extent is his current training regimen improving his cardiovascular fitness?
3. How can he modify his individual runs in order to optimize overall health benefits?

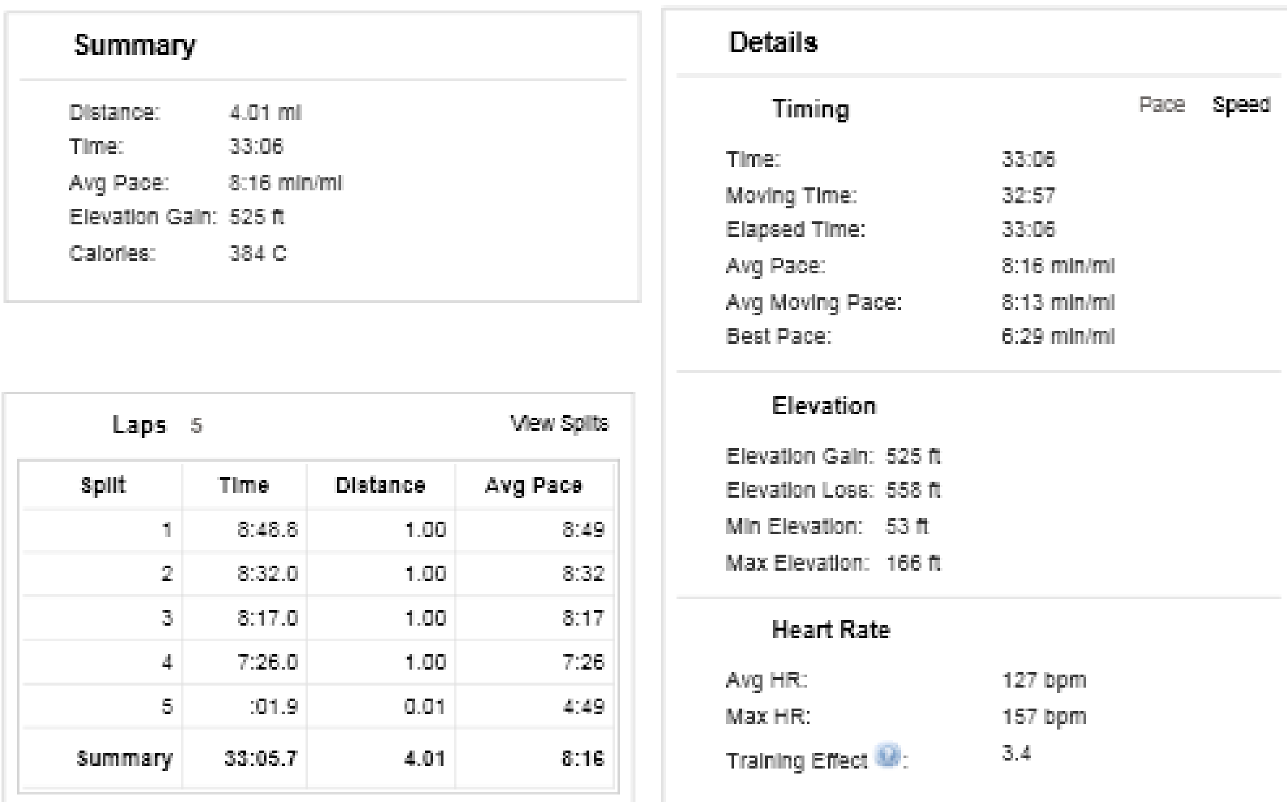
The remaining sections of this paper will address these questions while illustrating how the statistical analysis of this real life data incorporates the various recommendations of the GAISE report and others with regard to the use of technology to analyze real data in an active learning environment. Classroom implementation strategies will also be included.

2. Data Source & Description of Data

The data that is being analyzed here was collected with a Garmin Forerunner® 610 GPS watch, as shown in Figure 1. Figure 2 shows a typical display of the type of training data that is collected by this watch on any one run.

Figure 1. Garmin Forerunner® 610 GPS Watch

Figure 2. Typical Training Data Collected by Garmin Forerunner® 610 GPS Watch



Summary

Distance:	4.01 mi
Time:	33:06
Avg Pace:	8:16 min/mi
Elevation Gain:	525 ft
Calories:	384 C

Laps 5 View Splits

Split	Time	Distance	Avg Pace
1	8:48.8	1.00	8:49
2	8:32.0	1.00	8:32
3	8:17.0	1.00	8:17
4	7:26.0	1.00	7:26
5	:01.9	0.01	4:49
Summary	33:05.7	4.01	8:16

Details

Timing Pace Speed

Time:	33:06
Moving Time:	32:57
Elapsed Time:	33:06
Avg Pace:	8:16 min/mi
Avg Moving Pace:	8:13 min/mi
Best Pace:	6:29 min/mi

Elevation

Elevation Gain:	525 ft
Elevation Loss:	558 ft
Min Elevation:	53 ft
Max Elevation:	166 ft

Heart Rate

Avg HR:	127 bpm
Max HR:	157 bpm
Training Effect:	3.4

For purposes of this analysis, we have summarized in Table 1 some of the data from Figure 2 that will be analyzed or discussed in this paper. This data and associated codebook is also available at www.amstat.org/publications/jse/v22n2/Laumakis/Four-Mile_Run_Dataset.csv and www.amstat.org/publications/jse/v22n2/Laumakis/Four-Mile_Run_Dataset_Description.txt. The data associated with heart rate (HR) during the rest period shown in Table 1 was collected with the watch during a cool down period that began immediately after each run ended. The HR change data shown in Table 1 were computed by simply finding the difference in the HR at the beginning of the rest period to the one- and two- minute marks of the cool down period, respectively.

Table 1. Data Collected or Computed Associated with Kevin's Runs

Run	Time (min:sec)	Pace (min:sec)	Calories Burned	Training Effect	Max HR (BPM)	Avg HR (BPM)	Avg Speed (mph)	Max Speed (mph)	HR Rest (BPM)	HR Rest1 (BPM)	HR Rest2 (BPM)	HR Change 1 (BPM)	HR Change 2 (BPM)
1	36:09	9:00	319	3.0	150	123	6.7	7.8	126	85	78	41	48
2	33:06	8:16	384	3.4	157	127	7.3	9.3	150	102	94	48	56
3	32:59	8:15	398	3.5	156	132	7.3	9.1	146	99	93	47	53
4	32:17	8:04	359	3.7	159	130	7.5	8.9	152	106	94	46	58
5	34:53	8:43	366	3.6	168	135	6.9	8.3	140	77	88	63	52
6	32:04	8:01	388	3.6	159	138	7.5	9.7	121	80	72	41	49
7	32:29	8:07	411	3.8	163	142	7.5	9.2	135	85	90	50	45
8	31:19	7:50	423	3.9	164	144	7.7	9.5	140	90	89	50	51
9	35:06	8:46	373	4.1	160	140	6.9	8.6	137	102	100	35	37
10	33:39	8:25	418	3.5	150	136	7.1	8.6	122	84	84	38	38
11	32:37	8:09	446	3.8	159	140	7.4	8.8	133	94	95	39	38
12	32:06	8:02	400	3.9	162	144	7.5	9.1	139	90	89	49	50
13	35:34	8:53	347	3.3	172	148	6.8	7.9	145	112	108	33	37
14	36:14	9:04	334	3.5	161	140	7.0	8.3	145	89	93	56	52
15	34:48	8:42	378	1.1	143	103	6.9	8.8	123	102	94	21	29
16	31:43	7:56	368	3.5	147	131	7.6	9.5	117	89	80	28	37
17	35:46	8:57	320	3.2	158	136	6.7	8.2	145	101	100	44	45
18	34:19	8:35	326	3.5	160	136	7.0	7.9	144	97	94	47	50
19	34:17	8:34	302	3.8	173	153	7.0	8.1	143	86	76	57	67

3. Statistical Analysis

3.1 Training Effect (TE) Analysis

In order to assess how Kevin's current running program is improving his overall cardiorespiratory fitness, a detailed explanation of the TE variable is warranted. According to white papers entitled "EPOC Based Training Effect Assessment" and "Indirect EPOC Prediction Method Based on Heart Rate Measurement" (See http://www.firstbeat.fi/physiology/white-papers) published by Firstbeat Technologies Ltd., the company responsible for the software in the GPS watch, TE refers to training-induced development of fitness and performance. TE is measured on a scale from 1.0 to 5.0, with TE categories including Minor TE (1.0-1.9), Maintaining TE (2.0-2.9), Improving TE (3.0-3.9), Highly Improving TE (4.0-4.9), and Overreaching (5.0). In order to compute the TE for any single exercise routine, the peak EPOC (excess post-exercise oxygen consumption) achieved during exercise must be estimated. Typically, EPOC is a physiological measure of the amount of oxygen that is consumed in excess after exercise is complete. However, the GPS watch used here provides an estimate of the peak EPOC achieved during exercise, and this estimate is based on heart rate measurements recorded by the watch. By using the GPS watch, heart rate-based estimates for EPOC are evaluated at any given time during an exercise routine and higher EPOC estimates produce higher TE values. TE assessment provides key information on various exercise routines for a wide range of individuals from beginners to highly conditioned athletes.

The purpose of the analysis that follows in this subsection of the paper is to develop the best statistical model to predict TE based on the available heart rate data.

> *Alternative Applications: Although we focus on how TE depends on heart rate, others may choose to develop regression models that investigate other relationships based on other data in Table 1. For example, one may choose to investigate what relationship, if any, exists between the number of calories burned and the speed data provided.*

Using this best predictive model, an assessment of the effectiveness of Kevin's exercise program with respect to improved cardiorespiratory fitness can be done. Techniques used throughout this section of the paper are typically taught in both Statistics I & II courses at the college level and the software program used for the analysis here is JMP Pro 10. Appendix A illustrates the statistical analyses using R and Appendix B shows how SAS can be used to complete the analysis.

> *Helpful Hint: Before directing students to investigate the relationship between heart rate and TE, allow them to think about what variables provided in Table 1 may qualify as predictors for TE, along with their compelling reasons. This exercise provides students the important opportunity to consider possible relationships that make sense in the context of the story and data.*

Before we begin any analysis, an explanation of the low TE value for Run 15 is warranted. There are two compelling reasons as to why this particular TE value is low when compared to the TE values for the other runs. First, this particular run occurred around midnight, early in the month of July, whereas all the other runs were done during the day. During the month of July, the average daily high temperature was 89 degrees F and the average daily low temperature was 69 degrees F. So, the midnight run on that day was done at a temperature that was approximately 15 to 20 degrees F lower than the other runs that month. In their book, authors Benson and Connolly (2011, p. 31) address the impact of temperature on heart rate during exercise through the following excerpt, "An extremely important factor affecting exercise heart rate is temperature. Warmer temperatures cause the heart to beat faster and place considerable strain on the body. Simply put, when it is hot, the body must move more blood to the skin to cool it while also maintaining blood flow to the muscles. The only way to do both of these

things is to increase overall blood flow, which means that the heart must beat faster. Depending on how fit you are and how hot it is, this might mean a heart rate that is 20 to 40 bpm higher than normal." Indeed, comparing the Avg HR for Run 15 with the Avg HR values for the other runs, this appears to be the case. Since the GPS watch uses heart-rate based estimates for TE, we should expect to see a much lower TE for this run given the much lower Avg HR value.
The second reason why the TE is low on Run 15 is due to the extended rest between that run and the previous run. Whereas the other runs in the data set were done on two to three day's rest, Run 15 was done on seven day's rest. In their book, Benson and Connelly note that a longer recovery period between exercise routines results in a lower exercise heart rate of up to 5 to 10 BPM. So, like the lower Avg HR due to the lower temperature for Run 15, the longer break before that run contributed to a lower Avg HR and, in turn, to a lower TE value. As a result, although the low TE value for Run 15 may seem to be an outlier, on the contrary, it is to be expected given lower temperature conditions and extended rest.

As a first model, we consider whether the maximum heart rate (Max HR) alone is a good predictor of TE. Figure 3 shows the relevant output from JMP. The linear fit appears weak from the scatterplot and this weakness is supported by the relatively low R-Square value of 0.3373. Further, both the residual plot and the normal quantile plot for the residuals indicate a nonlinear relationship between Max HR and TE.

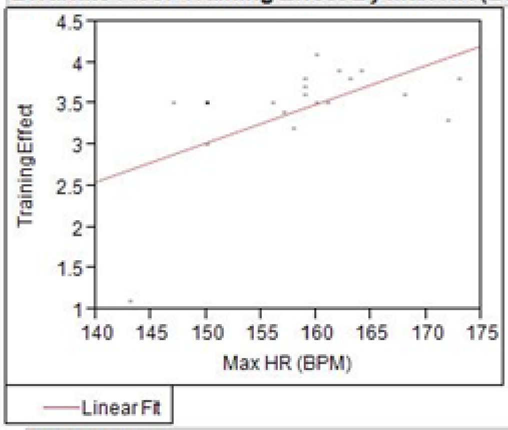
Examining the relationship between average heart rate (Avg HR) and TE produces the JMP output detailed in Figure 4. The scatterplot with the line of best fit appears to indicate a linear relationship between Avg HR and TE, and this relationship is a moderately strong relationship evidenced by the corresponding R-Square value of 0.6764. Also, at the 0.05 level of significance, the slope parameter is significantly different from zero.

> *Potential Pitfall: Upon initial inspection of the JMP output discussed above, students may erroneously conclude that a linear relationship exists between Avg HR and TE. However, the residual plot tells a different story here and demonstrates the importance of examining the residuals in regression analysis.*

However, examination of the residual plot and the normal quantile plot for the residuals leads to some suspicion over whether the relationship is indeed linear.

Figure 3. JMP Output for TE vs. Max HR

Bivariate Fit of Training Effect By Max HR (BPM)



Linear Fit

Training Effect = -4.049619 + 0.0472171*Max HR (BPM)

Summary of Fit

RSquare	0.33725
RSquare Adj	0.298265
Root Mean Square Error	0.527088
Mean of Response	3.457895
Observations (or Sum Wgts)	19

Analysis of Variance

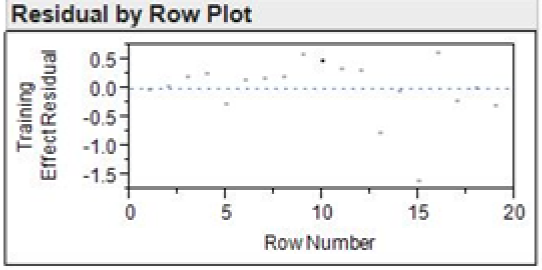
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.4033488	2.40335	8.6507
Error	17	4.7229670	0.27782	Prob > F
C. Total	18	7.1263158		0.0091*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-4.049619	2.555392	-1.58	0.1315
Max HR (BPM)	0.0472171	0.016054	2.94	0.0091*

Diagnostics Plots

Residual by Row Plot



Residual Normal Quantile Plot

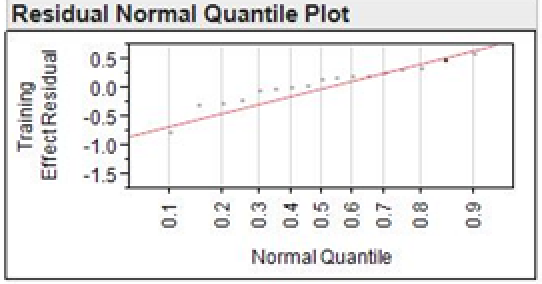
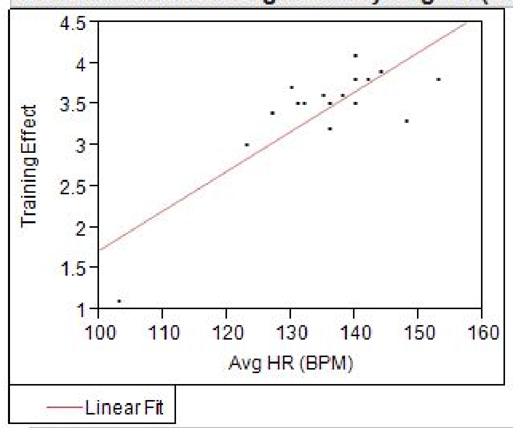


Figure 4. JMP Output for TE vs. Avg HR

Bivariate Fit of Training Effect By Avg HR (BPM)



Linear Fit

Training Effect = -3.099011 + 0.0483247*Avg HR (BPM)

Summary of Fit

RSquare	0.676404
RSquare Adj	0.657369
Root Mean Square Error	0.368307
Mean of Response	3.457895
Observations (or Sum Wgts)	19

Analysis of Variance

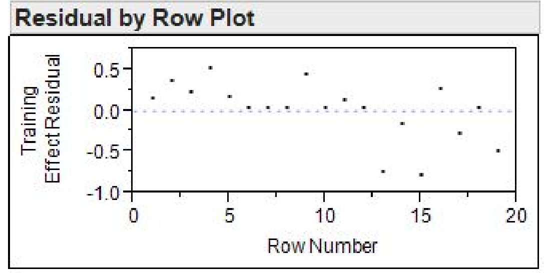
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.8202665	4.82027	35.5346
Error	17	2.3060493	0.13565	Prob > F
C. Total	18	7.1263158		<.0001*

Parameter Estimates

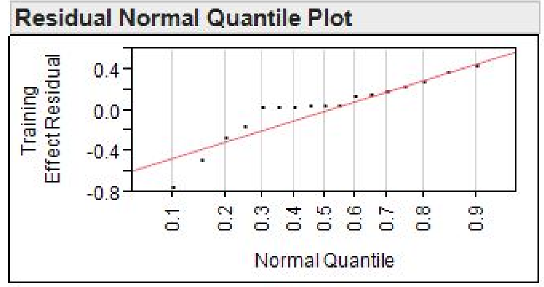
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3.099011	1.103191	-2.81	0.0121*
Avg HR (BPM)	0.0483247	0.008107	5.96	<.0001*

Diagnostics Plots

Residual by Row Plot

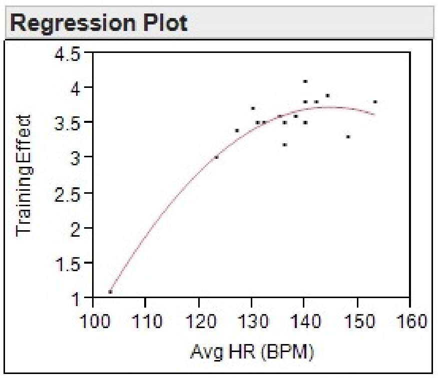


Residual Normal Quantile Plot



Since the relationship between Avg HR and TE appears nonlinear, we investigate a polynomial model. Figure 5 displays the pertinent output from JMP. The second-order polynomial appears to fit the data nicely and this fit is supported by the R-Square value of 0.8893. Further, the model parameters are highly significant given the reported *p*-values, and the lack of pattern in the residual plot implies that this model can be used to predict TE from Avg HR.

Figure 5. JMP output for Polynomial Model of TE vs. Avg HR



Summary of Fit

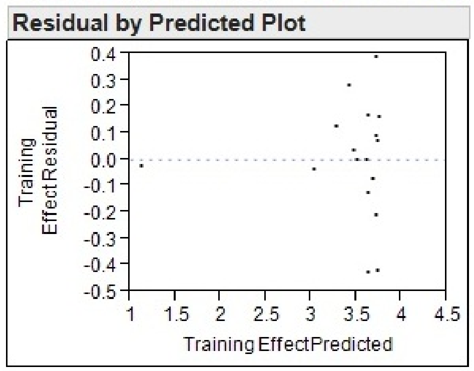
RSquare	0.889251
RSquare Adj	0.875408
Root Mean Square Error	0.222097
Mean of Response	3.457895
Observations (or Sum Wgts)	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	6.3370848	3.16854	64.2355
Error	16	0.7892310	0.04933	Prob > F
C. Total	18	7.1263158		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-27.90515	4.522561	-6.17	<.0001*
Avg HR (BPM)	0.4378422	0.070413	6.22	<.0001*
Avg HR (BPM)*Avg HR (BPM)	-0.001514	0.000273	-5.55	<.0001*



The parameters for this polynomial model are obtained from Figure 5, and the resulting estimating equation for TE is shown below:

$$\widehat{\text{TE}} = -27.9052 + 0.4378 * \text{Avg HR} - 0.0015 * \text{Avg HR}^2$$

This estimating equation for TE can now be used to determine the value of the Avg HR that will maximize TE.

> *Helpful Hint: For students who have had calculus, this is a typical optimization problem for which they may use the derivative of this equation to maximize TE. For students without a calculus background, graphing calculators provide a way to estimate the optimum value for Avg HR in order to maximize TE by use of numerical approximation.*

By taking the derivative of $\widehat{\text{TE}}$ with respect to Avg HR and setting the result equal to zero, we compute an Avg HR value of approximately 146 BPM in order to maximize TE. This result can be confirmed with a graphing calculator. The estimated maximum value for TE based on an Avg HR of 146 is computed directly to be 4.04.

The polynomial model is useful for determining the average heart rate that Kevin should try to achieve in order to maximize his training effectiveness. However, since it does not include any dependence on his maximum heart, we now explore some multiple regression models that investigate the dependence of TE on both Avg HR and Max HR. Figure 6 shows the multiple regression output from JMP using both Max HR and Avg HR as predictors of TE. Although 70.19% of the variability in TE is explained by this model as evidenced by the R-Square value, both the intercept and the Max HR parameters are highly insignificant.

> *Helpful Hint: Before directing students to investigate a multiple regression model that incorporates an interaction term, ask them to consider whether or not it would seem reasonable to conclude that Avg HR and Max HR do interact, along with their compelling arguments.*

In order to investigate any possible interaction between Max HR and Avg HR, Figure 7 shows the output from JMP for a multiple regression model that includes the interaction term Max HR*Avg HR. Not only are all terms in the model highly significant as evidenced by their associated *t* Ratios and *p*-values, but also approximately 90% of the variability in TE is explained by this model.

Figure 6. JMP Output TE vs. Max HR & Avg HR Model

Response Training Effect

Summary of Fit

RSquare	0.701921
RSquare Adj	0.664661
Root Mean Square Error	0.364366
Mean of Response	3.457895
Observations (or Sum Wgts)	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	5.0021100	2.50106	18.8385
Error	16	2.1242058	0.13276	Prob > F
C. Total	18	7.1263158		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.320275	1.871116	-0.71	0.4906
Max HR (BPM)	-0.022572	0.019287	-1.17	0.2590
Avg HR (BPM)	0.061666	0.013938	4.42	0.0004*

Figure 7. JMP Output for Interaction Model

Response Training Effect

Summary of Fit

RSquare	0.899517
RSquare Adj	0.87942
Root Mean Square Error	0.218491
Mean of Response	3.457895
Observations (or Sum Wgts)	19

Analysis of Variance

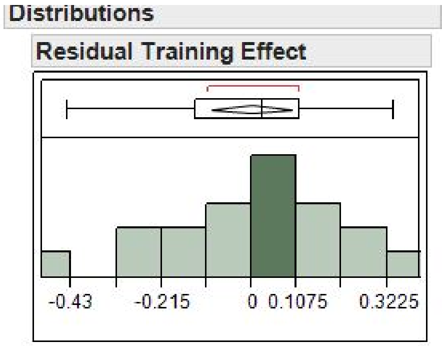
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	6.4102403	2.13675	44.7595
Error	15	0.7160755	0.04774	Prob > F
C. Total	18	7.1263158		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-57.08095	10.32805	-5.53	<.0001*
Max HR (BPM)	0.352678	0.070054	5.03	0.0001*
Avg HR (BPM)	0.4495128	0.0719	6.25	<.0001*
Max HR*Avg HR	-0.002612	0.000481	-5.43	<.0001*

In order to verify the assumptions required for this model to be valid, Figure 8 and Figure 9 are generated using JMP. Figure 8 shows that the distribution of the residuals for this model is approximately normal, as evidenced by the unimodal histogram centered approximately at zero with fairly symmetric tails. Figure 9 is the associated residual plot for this interaction model that shows uniform variance about the zero residual line and no apparent pattern in the plot. As such, this interaction model is valid and can be used to effectively predict TE from both Avg HR and Max HR.

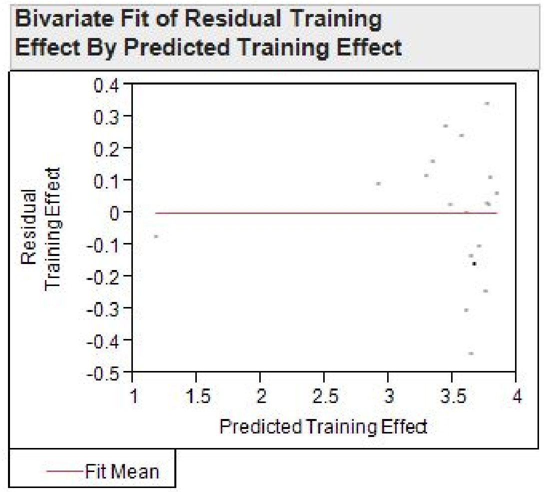
Figure 8. JMP Residuals Distribution for Interaction Model



100.0%	maximum	0.34013
99.5%		0.34013
97.5%		0.34013
90.0%		0.26835
75.0%	quartile	0.11253
50.0%	median	0.02706
25.0%	quartile	-0.1338
10.0%		-0.3044
2.5%		-0.439
0.5%		-0.439
0.0%	mir	

Mean	[illegible]
Std Dev	0.1994542
Std Err Mean	0.0457579
Upper 95% Mean	0.0961338
Lower 95% Mean	-0.096134
N	19

Figure 9. JMP Residual Plot for Interaction Model



Using this interaction model, we can now investigate various combinations of Avg HR and Max HR in an attempt to be able to tell Kevin how to maximize his TE on any given run. The parameters for the interaction model are obtained from Figure 7, and the resulting estimating equation for TE is shown below:

$$\widehat{\mathrm{TE}} = -57.0810 + 0.3527 * \mathrm{Max\ HR} + 0.4495 * \mathrm{Avg\ HR} - 0.0026 * \mathrm{Max\ HR} * \mathrm{Avg\ HR}$$

This estimating equation can be used to compute various estimates for TE based on different combinations of Avg HR and Max HR. If we set the Avg HR in equation above to the average value of 136 BPM over the nineteen 4-mile runs, we obtain the following:

$$\widehat{\mathrm{TE}} = 4.0510 - 0.0009 * \mathrm{Max\ HR}$$

In order to maximize the predicted TE using this new equation, Kevin should try to run so that his HR is a constant value of 136 BPM, which would result in an estimated TE of 3.93. However, given that it is not too realistic for him to be able to maintain a constant HR during the entire run, Table 2 shows estimates for TE based on various combinations of Avg HR and Max HR.

Table 2. Estimated TE Values Based on Various Combinations of Avg HR and Max HR

Avg HR	Max HR	TE
136	159	3.91
136	173	3.90
140	160	4.04
144	164	4.09
144	160	4.18
153	173	3.89

As shown in Table 2, the estimated TE values range from 3.89 to 4.18 for the various combinations of Avg HR and Max HR. Based on the results displayed in Table 2, Kevin should try to run so that his Avg HR is close to 144 BPM and his Max HR is close to 160 BPM in order to maximize his TE. If he can get close to these target values, his estimated TE will exceed 4.0, which will be in the "Highly Improving" range of training effectiveness and will serve to improve his overall cardiorespiratory fitness.

> *Potential Pitfall: Students with a background in multivariable calculus may attempt to take partial derivatives of $\widehat{TE}$ with respect to Max HR and Avg HR and set these forms equal to zero to determine their optimum values. This process will yield an optimum value of Max HR equaling 172.9 and Avg HR equaling 135.7, yielding a TE value of 3.90. So, while these values are candidates for optimum values of the multivariable function for $\widehat{TE}$, they are not the optimum values, clearly supported by examination of Table 2.*

3.2 Post-Exercise Heart Rate (PEHR) Analysis

PEHR recovery can be used as a powerful measure of cardiovascular fitness, and it is complementary to other traditional cardiovascular assessments including body fat, resting heart rate and blood pressure, and LDL, HDL, and total cholesterol (see Dimkpa 2009 and Shetler, Marcus, Froelicher, Vora, Kalisetti, Prakash, et al. 2001). Cardiovascular fitness refers to the combined efficiency of the heart, lungs, and vascular system in oxygen delivery to working muscles during prolonged exercise. PEHR recovery is a measure of the rate at which the heart rate decreases from the end of exercise to resting levels, and is an independent predictor of cardiovascular mortality in healthy adults (see Cole, Blackstone, Pashkow, Snader, and Lauer 1999). PEHR recovery to resting levels can take anywhere from one hour after moderate exercise up to four hours after long duration exercise routines, with the recovery in each case depending on the level of physical fitness of the exerciser. In assessing PEHR recovery, most investigators simply measure the change in heart rate from the end of exercise to either one or two minutes later. It has been shown that a decrease of 15-25 beats per minute (BPM) at the first minute of recovery is typical for a healthy person, whereas a first minute reduction of PEHR less than 12 BPM if recovery is active represents an unfavorable prognosis for cardiovascular mortality.

The purpose of the analysis that follows in this subsection of the paper is to assess Kevin's level of cardiovascular health as it relates to PEHR recovery. Table 1 shows his HR Change1 for each of his nineteen runs. This change data was computed by finding the difference between his heart rate at the beginning of his rest period (immediately after exercise) and his heart rate one minute later, for each of his nineteen runs. We want to determine if Kevin's true mean difference in heart rate immediately after exercise until one minute later indicates if he is heart-healthy. Normally we would test the following set of hypotheses:

$$H_0: \mu_1 - \mu_2 = 12$$
$$H_a: \mu_1 - \mu_2 > 12$$

where μ_1 is his true mean heart rate at the beginning of his rest period and μ_2 is his true mean heart rate one minute later.

> *Potential Pitfall: Students may treat this particular hypothesis test as one that involves independent samples, which is not the case. Since the data being analyzed is paired data for the same person, a matched-sample hypothesis test is warranted.*

However, since this test involves a matched-sample procedure, we may use JMP Pro 10 and simply define μ_d to be the true mean difference in Kevin's heart rate immediately after exercise until one minute later and test the following set of hypotheses:

$$H_0: \mu_d = 12$$
$$H_a: \mu_d > 12$$

Figure 10 displays the results of this hypothesis test with a test statistic of 13.6, and corresponding *p*-value less than 0.0001, indicating that Kevin's PEHR drop is significantly more than the 12 BPM associated with an unfavorable prognosis for cardiovascular mortality. Figure 10 also displays the normal quantile plot for the difference data, along with a histogram for the sample data. Based on these graphs, it appears that we may assume that the data come from an underlying population that is normally distributed. Further evidence that the sample data is drawn from an underlying normally distributed population is provided with the results of the Shapiro-Wilk Goodness-of-Fit test, also shown in Figure 10.

We may also investigate whether Kevin's PEHR drop during the first minute is significantly more than the upper end of the range for a typical healthy individual by testing the following set of hypotheses:

$$H_0: \mu_d = 25$$
$$H_a: \mu_d > 25$$

Figure 11 shows the results of this hypothesis test, producing a test statistic of 8.0482 and corresponding *p*-value once again less than 0.0001. Therefore, based on the preceding analysis, Kevin's PEHR drop during the first minute of recovery is not only significantly higher than those with an unfavorable prognosis for cardiovascular mortality, but is also significantly higher than the upper value of the range for a typical healthy person. Further, if Kevin is, in fact, not heart-healthy, then the probability of realizing his sample mean PEHR drop or greater purely by chance is given by the *p*-value and is less than 0.0001 for each significance test performed above. So, although it is possible to realize his sample mean PEHR drop or greater if he is not heart-healthy, it is not at all likely given the very small *p*-value in each case.

> *Alternate Application: We have conducted hypotheses tests to assess Kevin's level of cardiovascular fitness as it relates to PEHR recovery. As an alternative inferential assessment, one may choose to construct confidence intervals corresponding to varying levels of confidence (90%, 95%, 99%) to determine his cardiovascular fitness.*

Figure 10. JMP Output for PEHR Recovery Hypothesized at 12 BPM

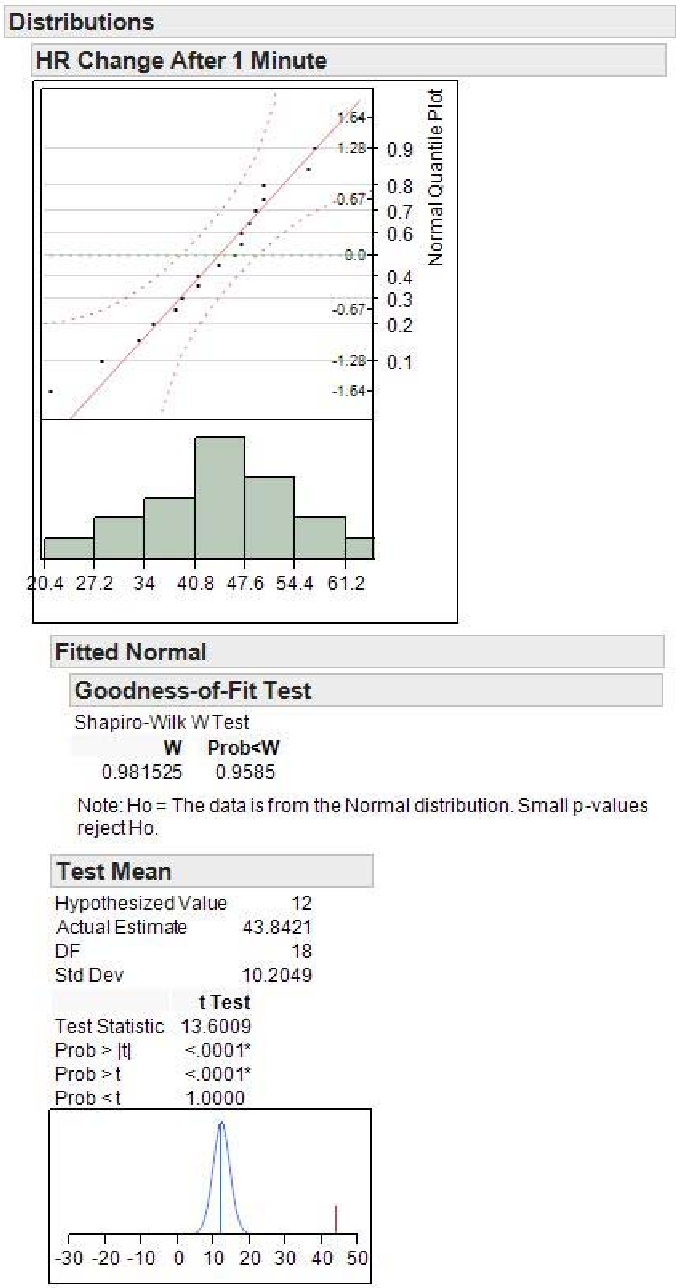
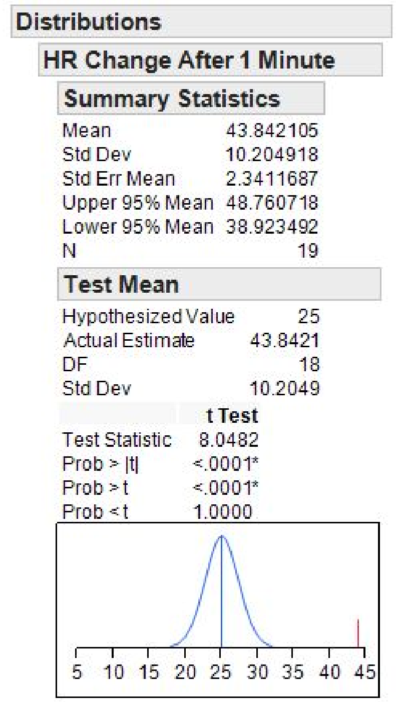


Figure 11. JMP Output for PEHR Recovery Hypothesized at 25 BPM



4. Classroom Implementation

The data shown in Table 1 can be used in college level Statistics I & II courses in a variety of ways. One way would be to introduce both the story and the data very early in the course, along with a detailed explanation of how the data were collected and what the various attributes measured mean. Then as the semester progresses, students can perform appropriate statistical analysis of the data in the classroom as they learn the pertinent statistical techniques. Using the story and the data in this manner serves to provide a course-long themed investigation to which students typically respond very well.

Another way to use the story and the data would be as a course project that the students could complete, either individually or in small groups. Such a project could be handled in two ways. One way would be to assign the full project early in the course, allowing the students to complete their analysis as they see fit, after relevant topics are covered in the classroom. Alternatively, the project could be introduced near the end of the course to serve as a capstone experience, perhaps even in lieu of a traditional final exam.

A third way to use the story and the data presented here would be as an end-of-course review. Although not specifically addressed in this paper, most of the topics traditionally covered in a Statistics I course could be reviewed with this data, including basic numerical and graphical statistics for one and two variable data sets, along with confidence intervals and hypothesis testing for unknown population means. Additionally, with some minor extension to the story, confidence intervals and hypothesis testing for unknown population proportions could also be included in the review. Similarly, the story and data given here lend themselves nicely to an end-of-course review for a typical Statistics II course, including such topics as two-population confidence interval and hypothesis testing, multiple regression techniques and inference related to these models' parameters, and even non-parametric methods.

Regardless of how the story and the data are used, most students are typically more motivated in their learning when they are able to apply what they are learning in the classroom to real life data. Moreover, if students become aware that they will be able to answer key questions related to the story and make sound recommendations based on their analysis, they tend to be more genuinely engaged in the learning process.

5. Summary & Conclusion

The main purpose of the analysis in this article is to answer the three questions posed in the first section of the paper with regard to Kevin's running program, given here again as follows:

1. How effective are his individual runs at improving his overall cardiorespiratory fitness?
2. To what extent is his current training regimen improving his cardiovascular fitness?
3. How can he modify his individual runs in order to optimize overall health benefits?

The analysis provided above will not only help Kevin to know how to modify his running routines to improve both his cardiorespiratory and cardiovascular fitness, but will also serve to motivate student learning in statistics.

Using statistical techniques typically taught in Statistics I & II courses at the college level, an analysis was performed to assess the effectiveness of Kevin's running program in improving the overall health of his heart, lungs, and vascular system. The results of the analysis provide Kevin an indication on how to adapt his runs so that he can consistently achieve a level of highly improving training effectiveness. Additionally, the analysis shows that one key measure of

Kevin's heart health, namely his PEHR recovery, is indicative of a person with excellent cardiovascular health.

Both the story and the data presented in this paper can be utilized in various ways in the college statistics classroom. Whether they are used as a course-long themed investigation, an individual or small-group project, or a course-end review in preparation for a final exam, the story and the data provide a compelling real life application of statistical methods. Through this story and the statistical analysis of its associated data sets, students will hopefully begin to recognize and appreciate how a regular exercise program can have a positive and measureable effect on one's health.

Appendix A

```
run <- read.csv("dataset.csv")

names(run)[5] <- "TE"
names(run)[6] <- "MaxHR"
names(run)[7] <- "AvgHR"
names(run)[13] <- "diffHR"  # drop in HR after 1 min of rest
```

Model #1: TE ~ MaxHR

```
mdl1 <- lm(TE ~ MaxHR, data = run)
par(mfrow = c(1, 1))
plot(TE ~ MaxHR, data = run)
abline(mdl1)
summary(mdl1)
##
## Call:
## lm(formula = TE ~ MaxHR, data = run)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.602 -0.132  0.142  0.271  0.609
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.0496     2.5554   -1.58   0.1315
## MaxHR         0.0472     0.0161    2.94   0.0091 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.527 on 17 degrees of freedom
## Multiple R-squared:  0.337,  Adjusted R-squared:  0.298
## F-statistic: 8.65 on 1 and 17 DF,  p-value: 0.00913
plot(residuals(mdl1))
abline(0, 0)
qqnorm(residuals(mdl1))
qqline(residuals(mdl1))
```

Model #2: TE ~ AvgHR

```
mdl2 <- lm(TE ~ AvgHR, data = run)
plot(TE ~ AvgHR, data = run)
abline(mdl2)
summary(mdl2)
##
## Call:
## lm(formula = TE ~ AvgHR, data = run)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7784 -0.0698  0.0402  0.1977  0.5168
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.09901    1.10319   -2.81    0.012 *
## AvgHR        0.04832    0.00811    5.96  1.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.368 on 17 degrees of freedom
## Multiple R-squared:  0.676,  Adjusted R-squared:  0.657
## F-statistic: 35.5 on 1 and 17 DF,  p-value: 1.55e-05
plot(residuals(mdl2))
abline(0, 0)
qqnorm(residuals(mdl2))
qqline(residuals(mdl2))
```

Model #3: TE ~ MaxHR + AvgHR

```
mdl3 <- lm(TE ~ MaxHR + AvgHR, data = run)
summary(mdl3)
##
## Call:
## lm(formula = TE ~ MaxHR + AvgHR, data = run)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7035 -0.1797  0.0429  0.1614  0.5926
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.3203     1.8711   -0.71  0.49058
## MaxHR        -0.0226     0.0193   -1.17  0.25900
## AvgHR         0.0617     0.0139    4.42  0.00043 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.364 on 16 degrees of freedom
## Multiple R-squared:  0.702,  Adjusted R-squared:  0.665
## F-statistic: 18.8 on 2 and 16 DF,  p-value: 6.23e-05
```

Model #4: TE ~ MaxHR*AvgHR

```
mdl4 <- lm(TE ~ MaxHR * AvgHR, data = run)
summary(mdl4)
```

```
##
## Call:
## lm(formula = TE ~ MaxHR * AvgHR, data = run)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4390 -0.1192  0.0271  0.1103  0.3401
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.71e+01   1.03e+01   -5.53  5.8e-05 ***
## MaxHR        3.53e-01   7.01e-02    5.03  0.00015 ***
## AvgHR        4.50e-01   7.19e-02    6.25  1.5e-05 ***
## MaxHR:AvgHR -2.61e-03   4.81e-04   -5.43  6.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.218 on 15 degrees of freedom
## Multiple R-squared:   0.9,   Adjusted R-squared:  0.879
## F-statistic: 44.8 on 3 and 15 DF,  p-value: 1.01e-07
hist(residuals(mdl4))
plot(fitted(mdl4), residuals(mdl4))
```

Informally estimating TE

```
mean(run$AvgHR)
## [1] 135.7
```

Test HR drop after 1 min

H-naught: diffHR = 12 H-alternative: diffHR > 12

```
mean(run$diffHR)
## [1] 43.84

xbar = mean(run$diffHR)
mu0 = 12
sigma = sd(run$diffHR)
n = nrow(run)
t = (xbar - mu0)/(sigma/sqrt(n))
t  # test statistic is 13.60094
## [1] 13.6

qt(0.9999, n - 1)  # critical t value at .0001 significant, 18 df is 4.648014
## [1] 4.648

shapiro.test(run$diffHR)
##
##  Shapiro-Wilk normality test
##
## data:  run$diffHR
## W = 0.9815, p-value = 0.9585
```

Test HR drop after 1 min

H-naught: diffHR = 25 H-alternative: diffHR > 25

```
mu0 = 25

t = (xbar - mu0)/(sigma/sqrt(n))
t  # test statistic is 8.048162
## [1] 8.048
```

Appendix B

Original SAS input file - www.amstat.org/publications/jse/v22n2/Laumakis/SASinput.sas

SAS output file follows.

1. Model result:

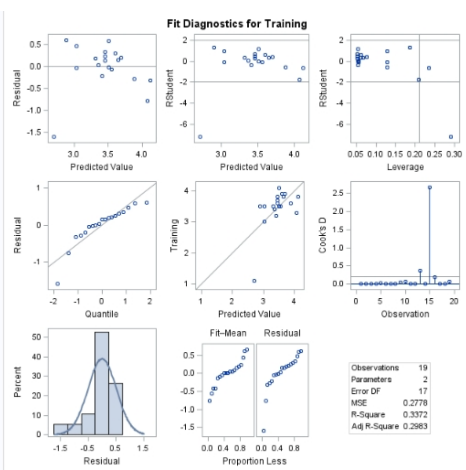
(1) Training effect as the response variable, maximum of Heart rate as the independent variable:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.40334879	2.40334879	8.65	0.0091
Error	17	4.72296700	0.27782159		
Corrected Total	18	7.12631579			

R-Square	Coeff Var	Root MSE	Training Mean
0.337250	15.24303	0.527088	3.457895

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-4.049619178	2.55539196	-1.58	0.1315
Max_HR	0.047217069	0.01605364	2.94	0.0091

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.84146	Pr < W	0.0049
Kolmogorov-Smirnov	D	0.196158	Pr > D	0.0524
Cramer-von Mises	W-Sq	0.145866	Pr > W-Sq	0.0241
Anderson-Darling	A-Sq	0.91712	Pr > A-Sq	0.0170



Based on the results, this model is not good for this problem. The residuals are not normal but independent in this model.

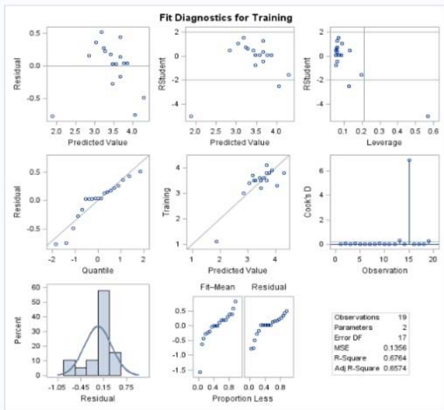
(2) Training effect as the response variable, average of Heart rate as the independent variable:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4.82026653	4.82026653	35.53	<.0001
Error	17	2.30604926	0.13564996		
Corrected Total	18	7.12631579			

R-Square	Coeff Var	Root MSE	Training Mean
0.676404	10.65119	0.368307	3.457895

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-3.099010658	1.10319129	-2.81	0.0121
Avg_HR	0.048324749	0.00810670	5.96	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.896345	Pr < W	0.0418
Kolmogorov-Smirnov	D	0.266735	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.151476	Pr > W-Sq	0.0213
Anderson-Darling	A-Sq	0.828411	Pr > A-Sq	0.0264



Based on all the results, this model is not good for this problem. The residuals are not normal but independent in this model.

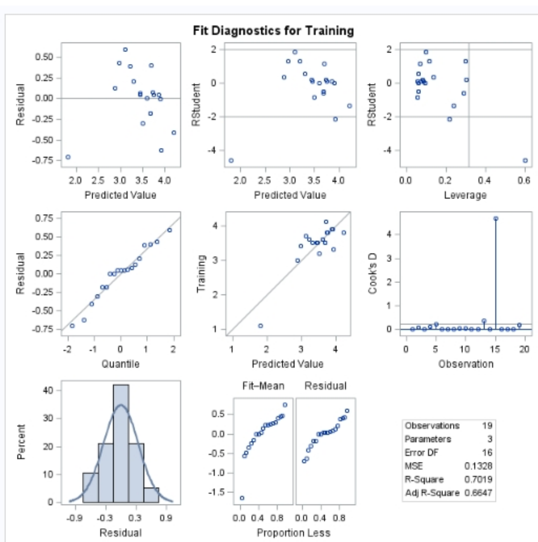
(3) Training effect as the response variable, average of Heart rate and maximum of Heart rate as the independent variables without interaction:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5.00211001	2.50105500	18.84	<.0001
Error	16	2.12420578	0.13276286		
Corrected Total	18	7.12631579			

R-Square	Coeff Var	Root MSE	Training Mean
0.701921	10.53723	0.364366	3.457895

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-1.320274945	1.87111568	-0.71	0.4906
Avg_HR	0.061665977	0.01393801	4.42	0.0004
Max_HR	-0.022571885	0.01928667	-1.17	0.2590

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.955994	Pr < W	0.4963
Kolmogorov-Smirnov	D	0.180749	Pr > D	0.0985
Cramer-von Mises	W-Sq	0.075472	Pr > W-Sq	0.2297
Anderson-Darling	A-Sq	0.403924	Pr > A-Sq	>0.2500



Based on all the results, this model is good for this problem. The R-square is high enough. The variables “Intercept” and “Max_Hr” are not significant in this model. But the residuals follow a normal distribution and appear independent.

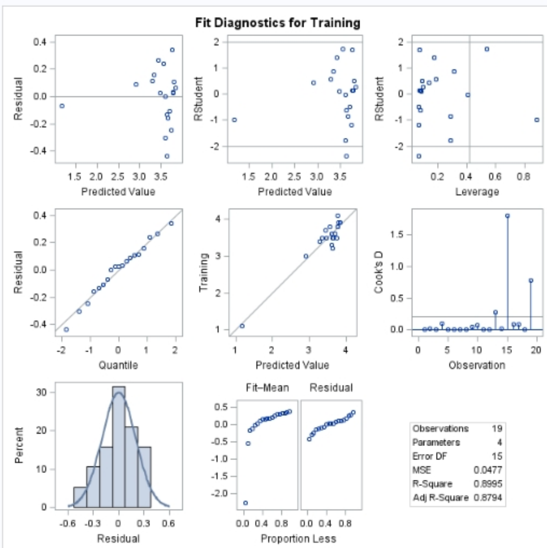
(4) Training effect as the response variable, average of Heart rate and maximum of Heart rate as the independent variables with interaction:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6.41024026	2.13674675	44.76	<.0001
Error	15	0.71607553	0.04773837		
Corrected Total	18	7.12631579			

R-Square	Coeff Var	Root MSE	Training Mean
0.899517	6.318617	0.218491	3.457895

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-57.08095350	10.32805244	-5.53	<.0001
Avg_HR	0.44951282	0.07189966	6.25	<.0001
Max_HR	0.35267801	0.07005407	5.03	0.0001
Avg_HR*Max_HR	-0.00261248	0.00048102	-5.43	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.979029	Pr < W	0.9305
Kolmogorov-Smirnov	D	0.127107	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.034224	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.197125	Pr > A-Sq	>0.2500



Based on the results, this model is really good for this problem. The R-square is really high. All variables are significant in this model, and the residuals follow a normal distribution and appear independent.

So, finally, we will choose the last model by considering all the factors in the model.

2. Hypothesis testing result:

(1) Test if the difference in the mean of Heart rate at the beginning and the mean of Heart rate after one minute is equal to 12 or greater than 12:

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	37.26044	Pr > |t|	<.0001
Sign	M	9.5	Pr >= |M|	<.0001
Signed Rank	S	95	Pr >= |S|	<.0001

In this table, we can see the difference is greater than 12.

(2) Test if the difference in the mean of Heart rate at the beginning and the mean of Heart rate after one minute is equal to 25 or greater than 25:

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	31.292	Pr > |t|	<.0001
Sign	M	9.5	Pr >= |M|	<.0001
Signed Rank	S	95	Pr >= |S|	<.0001

In this table, we can see the difference is greater than 25.

(3) Test if the difference in the mean of Heart rate at the beginning and the mean of Heart rate after one minute follows a normal distribution:

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.965861	Pr < W	0.6876
Kolmogorov-Smirnov	D	0.156802	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.060591	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.331965	Pr > A-Sq	>0.2500

In this table, we can see the difference follows a normal distribution.

Acknowledgement

We appreciate Suhwan Lee, University of Missouri, providing the SAS information in Appendix B.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., and Witmer, J. (2010), "Guidelines For Assessment and Instruction in Statistics Education: College Report." Technical report, American Statistical Association.

Benson, R. and Connelly, D. (2011), *Heart Rate Training*, Human Kinetics.

Cole, C.R., Blackstone, E.H., Pashkow, F.J., Snader, C.E., and Lauer, M.S. (1999), "Heart Rate Recovery Immediately After Exercise as a Predictor of Mortality," *New England Journal of Medicine*, 341, 1351-1357.

Dimkpa, U. (2009), "Post-Exercise Heart Rate Recovery: An Index of Cardiovascular Fitness," *Journal of Exercise Physiology*, 12(1), 10-22.

Firstbeat Technologies (2012), "EPOC Based Training Effect Assessment," White Paper available at http://www.firstbeat.fi/physiology/white-papers

Firstbeat Technologies (2007), "Indirect EPOC Prediction Method Based on Heart Rate Measurement," White Paper available at http://www.firstbeat.fi/physiology/white-papers

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. Y., and Ostrowski, E. (1996), *A Handbook of Small Data Sets*, London: Chapman and Hall.

Shetler, S., Marcus, R., Froelicher, V., Vora, S., Kalisetti, D., Prakash, M., et al. (2001), "Heart Rate Recovery: Validation and Methodological Issues," *Journal of the American College of Cardiology*, 38, 1980-1987.

Zieffler, A. and Karl, S. (eds.) (2010), "GAISE (Guidelines for Assessment and Instruction in Statistics Education) College Report," Available at http://www.amstat.org/education/gaise/

Paul J. Laumakis
Department of Mathematics
Rowan University
201 Mullica Hill Road
Glassboro, NJ 08028
laumakis@rowan.edu
