

Data Analysis and Commentary on Running Training Plan Effectiveness on The Islands

Brad Kreider

Abstract

The purpose of this research study is to measure the effect of different structured running plans on subjects' submaximal oxygen uptake (VO_2) within the virtual confines of The Islands. A total of 40 subjects were recruited and randomly assigned across four treatment groups, one of which was a control group. All prescribed running activity was conducted outdoors in order to maintain consistency of the training environment across treatment conditions. Statistical analysis consisted of one-way ANOVA and Tukey's test of multiple comparisons.

1. Background

The Islands are described as “a virtual human population that has been developed to support learning and teaching in experimental design, epidemiology and statistical reasoning” (University of Queensland). Developed by the University of Queensland, Australia, The Islands offers a virtual environment that is set up to enable realistic research studies on a virtual population and environment. This research study is taking place within the context of The Islands. All subjects are virtual and do not have any real world companions. Any conclusions obtained from this study have no merit or bearing outside of the Islands in the real world. Because the Islands is a virtual and entirely simulated population, there is no formal process for IRB review that is necessary for this study.

Context: A year on the Islands is 28 days long.

2. Motivations

Outside of my academic endeavors as a college student, I also participate in cross country and track & field at the NCAA Division III level. Thus, I am interested and curious in running and different training styles and metrics. These interests have motivated me to pursue this study on The Islands to study the effects of different running training plans on subjects' submaximal oxygen uptake (VO_2 max).

Question: How do different running training plans affect subjects' submaximal oxygen uptake (VO_2 max) on The Islands?

Expectations: Based upon my real world experience and knowledge of running and aerobic capacity, I would expect to see significant differences between the changes in VO_2 max among the treatment groups.

3. Study Overview

The study took place over 7 days in the real world, which corresponds to approximately ¼ year on the Islands.

3.1 Recruitment

40 participants were selected randomly from a randomly selected town on the Islands. (Nelson) Eligible participants were between the ages of 18–55, gender was not deemed to be an important factor for this study.

Homes in Nelson were selected randomly using a random number table. The first eligible participant from each home was asked for consent to be included in the study. If the subject declined to participate or if there were no eligible subjects, the house was passed over.

3.2 Experiment Design

Subjects were placed randomly into 1 of 4 treatment groups. A summary of treatment groups is listed below:

| Day | Control | Minimal | Moderate | Maximal |
|-----|---------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| 1 | Pre-test VO ₂ | Pre-test VO ₂ ; 1 km run | Pre-test VO ₂ ; 5 km run | Pre-test VO ₂ ; 5 km run |
| 2 | Rest | Rest | Rest | 5 km run |
| 3 | Rest | 1 km run | 5 km run | 5 km run |
| 4 | Rest | Rest | Rest | 5 km run |
| 5 | Rest | 1 km run | 5 km run | 5 km run |
| 6 | Rest | Rest | Rest | 5 km run |
| 7 | Post-test VO ₂ | 1 km run; Post-test VO ₂ | 5 km run; Post-test VO ₂ | 5 km run; Post-test VO ₂ |

3.3 Tracking

Participant info (name), Pre and post VO₂ measurements, and treatment record (checklist) were recorded in an excel spreadsheet. Please see treatment group 3, “Moderate Running Group” data sheet below.

| Treatment Group: Moderate Running Plan | | | | | | | | | | | |
|--|-----------------|----------|----------------|----------|----------------|----------|-----------------|--------------------------|---------------------------|--------------|----------------|
| Prescription: 5 km outdoors | | | | | | | | | | | |
| Run Compliance Tracker | | | | | | | | | | | |
| Subject Name | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 | Pre-test VO ₂ | Post-test VO ₂ | Change | Runs ✓ |
| (Islander) | Pre-test + Run | Rest | Run | Rest | Run | Rest | Run + Post-test | (mL/kg/min) | (mL/kg/min) | Post - Pre | of prescribed |
| Legend: | = Run completed | | Pre-test day | | Post-test day | | Rest day | | | | |
| Subjects | | | | | | | | | | | |
| Activity | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 | Pre-test VO ₂ | Post-test VO ₂ | Change | Runs ✓ |
| | Pre-test + Run | Rest | Run | Rest | Run | Rest | Run + Post-test | (mL/kg/min) | (mL/kg/min) | Post - Pre | of prescribed |
| Gersende Babu | X | — | X | — | X | — | X | 56.1 | 53.7 | -2.40 | 4 / 4 |
| Jacqueline Baumann | X | — | X | — | X | — | X | 38.9 | 36.7 | -2.20 | 4 / 4 |
| Alick Black | X | — | X | — | X | — | X | 66.7 | 65.8 | -0.90 | 4 / 4 |
| Yuka Regan | X | — | X | — | X | — | X | 42.8 | 39.7 | -3.10 | 4 / 4 |
| Kyosuke Koyama | X | — | X | — | X | — | X | 50.3 | 48 | -2.30 | 4 / 4 |
| Hiromori Hasegawa | X | — | X | — | X | — | X | 60.8 | 58.4 | -2.40 | 4 / 4 |
| Daniel Jensen | X | — | X | — | X | — | X | 79.8 | 77.2 | -2.60 | 4 / 4 |
| Rin Collins | X | — | X | — | X | — | X | 38.8 | 36.7 | -2.10 | 4 / 4 |
| Noel Sato | X | — | X | — | X | — | X | 46.5 | 44 | -2.50 | 4 / 4 |
| Arabelle Bager | X | — | X | — | X | — | X | 38.7 | 35.7 | -3.00 | 4 / 4 |
| Totals | 10 / 10 | — | 10 / 10 | — | 10 / 10 | — | 10 / 10 | 51.94 | 49.59 | -2.35 | 40 / 40 |

Moderate Running Group tracking sheet

Data can be obtained at <https://github.com/bradley-kreider/Data-Analysis-on-The-Islands/tree/master>

Note: On day 3 of the study, one of our subjects in the maximal training group ‘withdrew’ from the study. I received no additional information as to the reason for their withdrawal. Given that we were already 3 days into the study, we proceeded without replacement of another subject. Thus, in our analysis, the maximal training group has 9 participants instead of 10.

All statistical analysis that follows was conducted at the 95% confidence level.

3.4 VO₂ Max

“VO₂ max measures how much oxygen your body consumes while exercising” (Warner, 2024). Typically measured in ml/kg/minute, VO₂ max is essentially a measure of how aerobically fit an individual is. We will be using this metric to track the effectiveness of the treatment groups in this study, measured in ml/kg/minute.

To achieve this metric, we performed a test of submaximal oxygen uptake on our subjects. While this test is certainly not as precise as performing a full scale laboratory VO₂ maximum test, this provides a great starting point for which to measure the general change of aerobic capacity for the running training plans.

4. Analysis

To study the changes in VO₂ max from subject to subject, we are looking at (post–pre) (VO₂) for each subject. Below are the summary statistics for the 4 treatment groups.

Statistics

| Variable | N | N* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---------------------------|----|----|----------|----------|----------|---------|--------|--------|--------|---------|
| Control Group V02 Change | 10 | 0 | -0.08 | 0.203743 | 0.644291 | -0.8 | -0.625 | -0.25 | 0.675 | 0.9 |
| Minimal Group V02 Change | 10 | 0 | -1.59 | 0.189414 | 0.598981 | -2.6 | -2.05 | -1.5 | -1.175 | -0.7 |
| Moderate Group V02 Change | 10 | 0 | -2.35 | 0.190467 | 0.602310 | -3.1 | -2.7 | -2.4 | -2.175 | -0.9 |
| Maximal Group V02 Change | 9 | 0 | -1.31111 | 0.274593 | 0.823779 | -2.2 | -1.9 | -1.5 | -0.95 | 0.5 |

Summary statistics for the four treatment groups

For our primary analysis of the treatment groups, we used a one-way ANOVA test to test for differences in mean VO_2 change for each treatment group.

An ANOVA test was used because we have 4 independent groups (treatment groups) that we are studying that each have their own mean (post-pre VO_2). Thus, an ANOVA test is the perfect candidate for this analysis.

4.1 Conditions for One-Way ANOVA

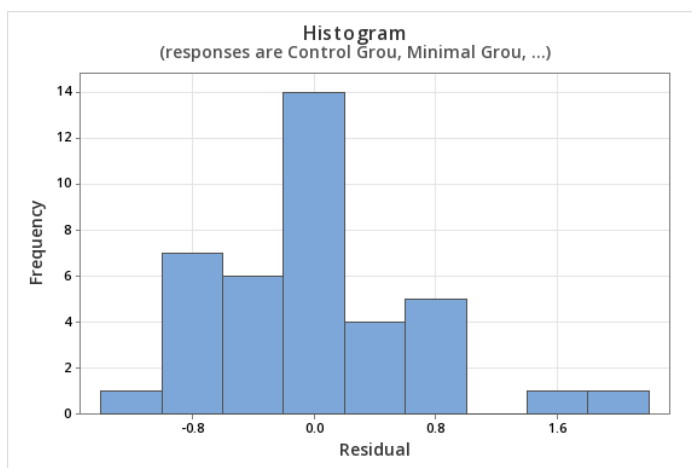
For this one way ANOVA test, we need the following conditions to be met:

Independence / Random

We need observations that are independent of one another in our study. In the recruitment of our study, we made use of random sampling to select participants as well as random assignment of treatment groups. Each subject is theoretically independent of one another, as we did not select more than 1 subject per household. The independence condition is satisfied for this study.

Normality

We need our subjects to be sourced from a normally distributed population.

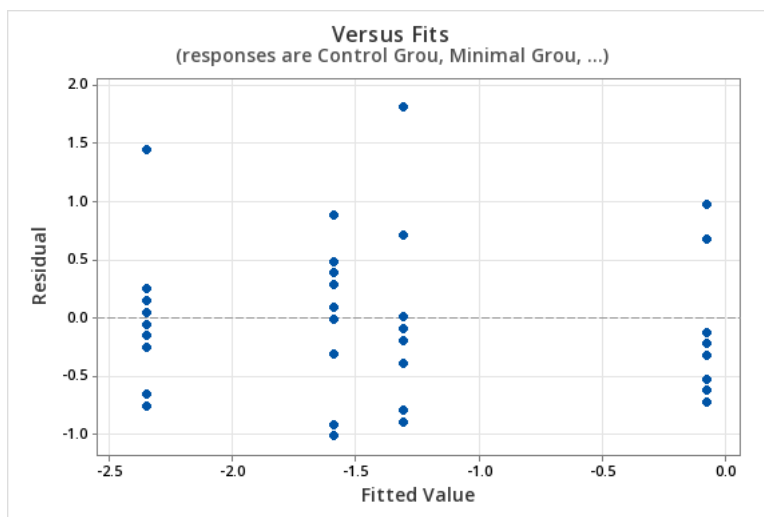


Histogram of residuals vs. frequency

Looking at our histogram of Residual vs Frequency, we can see that our data somewhat resembles a normal distribution, and appears to be unimodal. However, there appears to be more residual on the right-hand side and the distribution fit is far from perfect. Thus, we will assume the normality condition has been met and will proceed with caution.

Equal Variance

We need our subjects to be sourced from a population with equal variance.



Fitted values vs. residuals

Looking at the graph of our fitted value vs Residual, we can see that we have fairly scattered residuals for each of our treatment groups, with spread that is roughly equal across the groups. This spread is acceptable to assume equal variance.

Normality is questionable so we will proceed to the ANOVA analysis with caution.

4.2 ANOVA Analysis

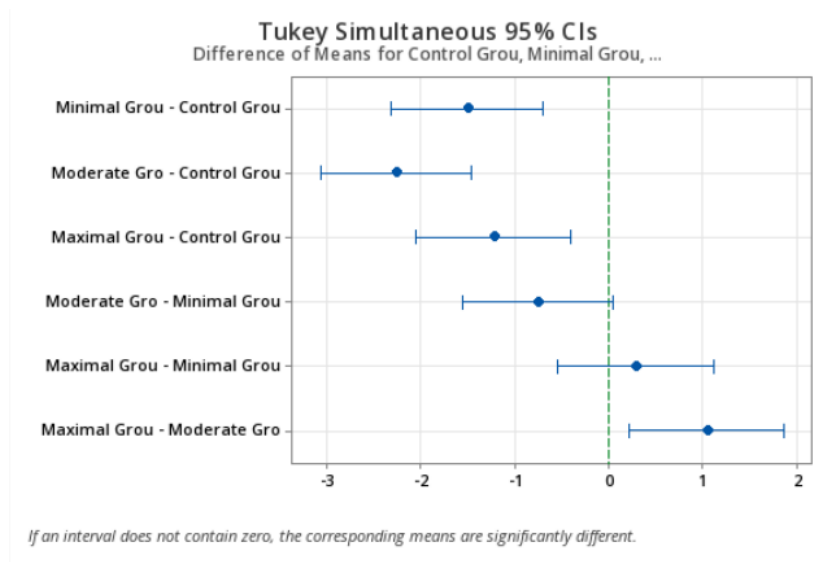
Our one-way ANOVA was significant with an F-statistic of 19.9 and a p-value of approximately 0.000. Thus, we have sufficient evidence to conclude that there exists some difference between the mean differences in VO_2 (post-pre) for all subjects. Our ANOVA test does not inform us as to where these differences come from, so we must turn to a test of Tukey's Multiple Comparisons to inform us as to where these differences occur.

Our ANOVA test resulted in a R^2 of 63.04%, meaning that our model explains approximately 63% of the variation of VO_2 change across the treatment groups. Thus, we can conclude that the running treatment plans do indeed account for a significant amount of the differences in VO_2 .

4.3 Tukey's Multiple Comparisons

Given the results of our significant ANOVA table, we will run a test of Tukey's Multiple Comparisons (MC) to locate where exactly these differences occur.

Looking at the output of our MC test, we gain further insights into the data. It is clear that treatment group #1 (control) displays a significant difference from all 3 treatment groups, and is the only group whose 95% confidence interval includes 0.



Tukey's 95% Simultaneous Confidence Intervals

For the control group, our confidence interval extends on both positive and negative sides of 0. Having a mean difference very close to 0 means that the control group truly has no meaningful change in VO_2 , as a 95% confidence interval for this group includes the value 0. With a mean of -0.080, this group is significantly different from the other 3 treatment groups that have means in the range (-1.311, -2.350).

Tukey's MC also informs us that the minimal and maximal training plan groups are indistinguishable at the 95% level, along with the moderate and minimal groups. Only the moderate and maximal groups differ significantly from each other ($\mu_4 > \mu_3$).

5. Conclusions

Given my interests in running that motivated this study, I would have expected all of the treatment groups, with the exception of potentially the control group, to have exhibited positive mean differences in VO_2 (post-pre). In the real world, a running training plan would typically result in a slow increase in submaximal oxygen uptake. Thus, in the Islands, I would have expected to see mean positive differences in the minimal, moderate, and maximal training plan groups, with the maximal training group being the largest difference (increase), the moderate being in the middle, and the minimal being the smallest increase. In terms of the control group, I would have expected to have seen VO_2 mean difference be roughly around 0, or potentially slightly negative or positive, but very close to 0 change.

In reality, these expectations were far off of what really took place on the Islands. The control group somewhat met my expectations, resulting in a mean difference of -0.080. Apart from this, my expectations were largely incorrect. Across the various treatment groups that received a running training plan, we saw across the board a decrease (negative mean difference post-pre) in VO_2 max after prescribing a running

training plan. Here lies the extent of our statistical analysis. I cannot statistically derive any further conclusions as to why this is the case.

6. Possible Explanations

While I cannot show statistical evidence of any reason for the decrease in mean VO_2 (post-pre) for all the treatment groups, there are some potential confounding factors that could account for this outcome for the Islanders.

1. **Fatigue.** Subjects without a normal habit of running that were introduced into one of the running training plans may have simply become fatigued as a result of the training plan, resulting in a decrease in aerobic capacity at the end of the study.
2. **Factors of the Islands.** Of course, it is also entirely possible that the Islands are simply not set up to account for running being a contributing factor to a subject's VO_2 max. If this were to be the case, no matter how much the subjects ran, their VO_2 max would not be affected.

Appendix A: Formal 6-Step Hypothesis Test

Hypothesis: In this ANOVA Analysis, the null hypothesis assumes that the mean differences in VO₂ (post–pre) for all treatment groups are equal ($\mu_i = \mu_j \forall i,j$). The alternative hypothesis assumes that the mean differences in VO₂ (post–pre) are not equal in some way ($\exists i,j : \mu_i \neq \mu_j$). Where μ_1 is the difference in VO₂ (post–pre) for all islanders who would not be on a running treatment plan, μ_2 is the difference in VO₂ (post–pre) for all islanders who would be on a minimal running treatment plan, μ_3 is the difference in VO₂ (post–pre) for all islanders who would be on a moderate running treatment plan, μ_4 is the difference in VO₂ (post–pre) for all islanders who would be on a maximal running treatment plan.

Conditions: The conditions for our ANOVA (Independence, randomness, normality, equal variance) have been satisfied, we are proceeding with caution with the normality condition. See above.

Rejection Region: We will reject H₀ if TS > 2.87

Test Statistic: TS = 19.90

P-value: $p \approx 0.000$

Conclusion: We have enough evidence ($p \approx 0.000$) to conclude that the 4 treatments (running training plans) differ with respect to their mean difference in VO₂ (post–pre) for all treatments.

Appendix B: Formal Tukey's MC Conclusions

$\mu_1 > \mu_2, \mu_3, \mu_4 ; \mu_4 > \mu_3$

Grouping Information Using the Tukey Method and 95% Confidence

| Factor | N | Mean | Grouping |
|---------------------------|----|--------|----------|
| Control Group V02 Change | 10 | -0.080 | A |
| Maximal Group V02 Change | 9 | -1.311 | B |
| Minimal Group V02 Change | 10 | -1.590 | B C |
| Moderate Group V02 Change | 10 | -2.350 | C |

Means that do not share a letter are significantly different.

Tukey's multiple comparisons output.

Appendix C: Additional Minitab Output

Factor Information

| Factor | Levels | Values |
|--------|--------|---|
| Factor | 4 | Control Group V02 Change, Minimal Group V02 Change, Moderate Group V02 Change, Maximal Group V02 Change |

Factor information.

Method

| | |
|------------------------|-------------------------|
| Null hypothesis | All means are equal |
| Alternative hypothesis | Not all means are equal |
| Significance level | $\alpha = 0.05$ |

Equal variances were assumed for the analysis.

ANOVA Setup

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|--------|----|--------|--------|---------|---------|
| Factor | 3 | 26.71 | 8.9026 | 19.90 | 0.000 |
| Error | 35 | 15.66 | 0.4474 | | |
| Total | 38 | 42.37 | | | |

One-way ANOVA Output

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|----------|--------|-----------|------------|
| 0.668877 | 63.04% | 59.87% | 53.97% |

Model Summary

Means

| Factor | N | Mean | StDev | 95% CI |
|---------------------------|----|--------|-------|------------------|
| Control Group V02 Change | 10 | -0.080 | 0.644 | (-0.509, 0.349) |
| Minimal Group V02 Change | 10 | -1.590 | 0.599 | (-2.019, -1.161) |
| Moderate Group V02 Change | 10 | -2.350 | 0.602 | (-2.779, -1.921) |
| Maximal Group V02 Change | 9 | -1.311 | 0.824 | (-1.764, -0.858) |

Pooled StDev = 0.668877

Group Means

References

Minitab, LLC. (2024). *Minitab* (Version 21) [Software]. <https://www.minitab.com>

University of Queensland. (n.d.). *The Islands: A virtual environment for teaching epidemiology and statistics* [Software]. <https://islands.smp.uq.edu.au/>

Warner, L. (2024, July 8). *VO₂ max: What is it and how can you improve it?* Harvard Health. <https://www.health.harvard.edu/healthy-aging-and-longevity/vo2-max-what-is-it-and-how-can-you-improve-it>