Introduction
Mental health is a crucial area of emphasis for improving global quality of life, and rapid-effect interventions are a key component of this goal. One often suggested such intervention is exercise, which is said to be good for both body and mind. We investigate if this relationship is truly causal using a Randomized Complete Block Design (RCBD); in particular, we are concerned with testing multiple levels of exercise intensity across varying age brackets. By considering multiple levels of intensity, we are able to determine if more intense exercise leads to greater effects on mood, or if any exercise at all is sufficient.

Our research question aims to investigate how varying intensities of outdoor exercise affect short-term self-reported negative mood. Using a random sample of 108 participants blocked by age bracket, we assign three levels of exercise intensity for identical periods of time. We measure the difference in self-reported negative affect from before and after in order to test if different intensities of outdoor exercise produce a change in mood.

We used the simulation world *The Islands* as the source for participants, and it served as the setting for the research. The world consists of three islands, each with multiple villages and thousands of simulated individuals. Users are able to gain consent from the islanders in order to assign interventions, and conduct surveys. There are multiple interventions available for exercise, including three different intensities of outside exercises: “Relaxing Walk Outdoors”, “Brisk Walk Outdoors”, and “Run Outdoors” all for 30 minutes in length. Moreover, the islanders are able to respond to questions from the short version of the *Profile of Mood States* mood scale via a survey function, allowing us to access self-reported mood information.

Source: *The Islands* [https://islands.smp.uq.edu.au]

Design of the Experiment
We used an RCBD method in order to carry out this analysis. This methodology allows us to test multiple levels of a treatment factor, which is important in order to fully understand the effects of varying intensities of exercise on negative mood. RCBD also allows for the inclusion of a nuisance factor to consider, which is crucial for accounting for natural differences in the population. In this study, we blocked by age, which lets us interpret the results separately for people of different physical maturities, which might otherwise affect the impact of exercise. The final design was an RCBD, with one treatment factor of **exercise type** and one nuisance factor of **age**, both with three levels. The three levels of the treatment factor, exercise type, are: “Relaxing Walk Outdoors 30 mins” (1), “Brisk Walk Outdoors 30 mins” (2), and “Run Outdoors 30 mins” (3). The nuisance factor, age, has the following three levels: 19-24 (1), 25-44 (2), and 45-64 (3). The outcome variable measured was **difference in negative mood**, which was calculated to be the post-intervention overall negative mood subtracted from the pre-intervention overall negative mood (i.e., pre - post). This means that an improvement in mood, or a decrease in overall negative mood, after the intervention would result in a *positive* outcome variable.

Using birth records from the six villages on the island of Ironbard (records are found in the Town Hall of each respective village), we used a random number generator to randomly sample 36 participants from each of the three age brackets across those born in Ironbard. Sampling in this way allowed us to access the full population of those born on Ironbard, in addition to their ages in order to facilitate blocking. As such, we were able to ensure proper randomization and replication of our sample for the most internally valid experimental results. If a selected participant was not alive or declined to participate, we continued using the random number generator to select additional participants, repeating until we had a total of 108 consenting participants. Within each age bracket, we again used a random number generator to randomly assign the three treatments equally between the 36 participants such that for each age bracket, there are 12 participants assigned to each treatment level (12 each to relaxing walk, brisk walk, and run).

Using an RCBD design, the statistical model for our data is:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

$i = 1, 2, 3$ (i^{th} treatment level/exercise type)

$j = 1, 2, 3$ (j^{th} block/age bracket)

$y_{ij} =$ observation from the ith level of exercise intensity and jth age bracket

$\mu =$ overall mean difference in mood

$\tau_i =$ effect of ith level of exercise intensity

$\beta_j =$ effect of jth age bracket

$\varepsilon =$ random error term; we assume that $\epsilon \sim \mathcal{N}(0, \sigma^2)$

To control for variation in time of day, all experiments were conducted in the afternoon beginning at 1pm. Using the survey feature of *The Islands* to ask islanders questions, we began by first asking each participant nine questions indicating subjective levels of three negative mood states (anger, depression, tension) from "A Shortened Version of the Profile of Mood States" (Shacham, 1983). Each question was scored with an integer between 0 and 4, and the following rule was used to translate participants' verbal answers to numerical code: 0 - "not at all", 1 - "a little", 2 - "moderately", 3 - "quite a bit", 4 - "extremely". The maximum possible survey score, 36, represents the highest possible level of negative mood reported by the subject across all three mood states, and a score of 0 represents the lowest possible level of negative mood. Hereafter, we refer to "overall negative mood" as the degree of combined anger, depression, and tension reported by the subject. Hence, this score out of 36 represents their initial, pre-intervention level of overall negative mood.

Anger	Depression	Tension
How angry are you feeling?	How discouraged are you feeling?	How anxious are you feeling?
How annoyed are you feeling?	How unhappy are you feeling?	How restless are you feeling?
How bitter are you feeling?	How worthless are you feeling?	How tense are you feeling?

Table 1. Nine selected questions from "Profile of Negative Mood States" corresponding to three negative moods.

Each participant then performed their assigned treatment, either running, brisk walking, or relaxing walk outdoors. All interventions were 30 minutes in length. After exercising, participants were immediately given the nine question survey again, which yielded another score out of 36, representing their post-exercise level of overall negative mood.

To arrive at our final measure of mood (our response variable), we then subtracted the post-exercise survey score from the initial survey score to yield a single number which represents *change in overall negative mood* due to exercising:

$$\text{change in overall negative mood} = \text{Pre-exercise score} - \text{Post-exercise score}$$

Therefore, a *positive* response variable represents improvement in negative mood (lower levels of anger, tension, and/or depression relative to baseline) after exercising. A *negative* response variable represents a

worsening of mood (higher levels of anger, tension, and/or depression relative to baseline) after exercising. The magnitude of the response variable, which has a theoretical range from -36 to 36, represents the degree of change in mood due to exercise.

```
##      Balanced one-way analysis of variance power calculation
##
##               k = 3
##               n = 31.60263
##               f = 0.3713907
##       sig.level = 0.05
##           power = 0.9
##
## NOTE: n is number in each group
```

Figure 1. R output from power calculations for balanced one-way ANOVA.

We then used power calculations to check that our sample size was appropriate. For effect size, we used the ratio between the maximum difference between treatment means we wished to detect (1 unit), and σ (approximated by $\sqrt{MSE} \approx 2.69258$). We found that at least 32 subjects in each of the 3 treatment groups would give us a power of at least 0.9, meaning that the experiment has at least a 90% probability of detecting a statistically significant difference of 1-unit between treatment means. Since we conducted our experiment with 36 subjects in each treatment group, and the treatment means all differ by more than 1 unit (see below), we deemed our sample size to be adequate.

Results and Interpretation

Our null hypothesis, H_0, is that the three levels of exercise intensity have the same effect on negative mood ($\tau_1 = \tau_2 = \tau_3 = 0$), meaning that the treatment effect does not contribute significantly to our observed difference in overall negative mood. Our alternative hypothesis H_A is that $\tau_i \neq 0$ for at least one *i*, meaning that at least one level of exercise intensity has a different effect on overall negative mood than the others.

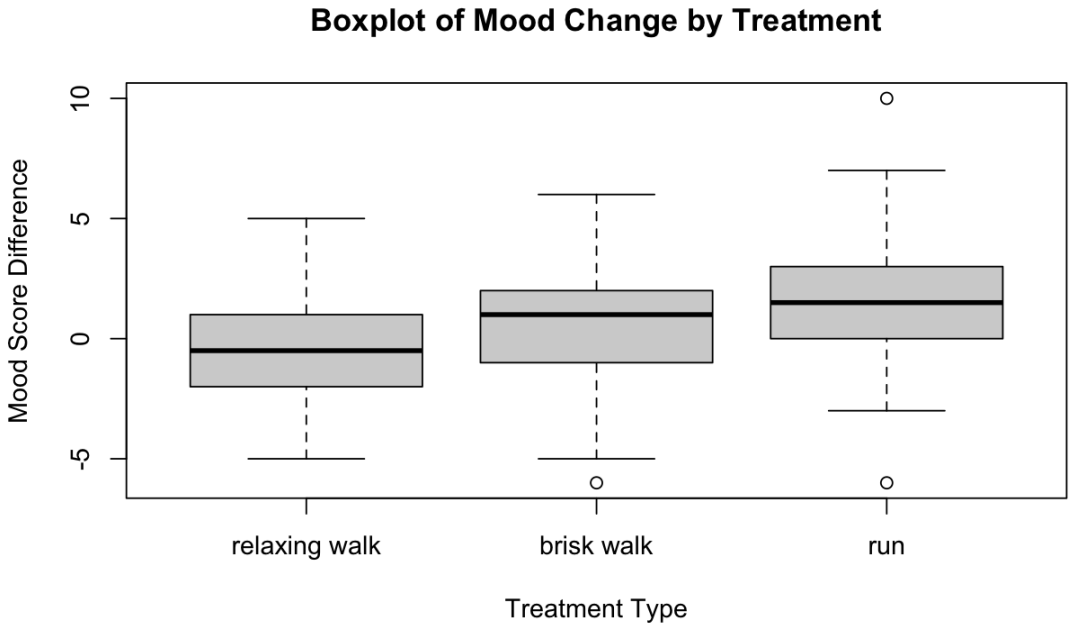

Figure 2. Boxplot showing difference in overall negative mood as a result of different exercise types/intensities.

Treatment	**Mean** (Change in overall negative mood)	**Standard deviation**
Relaxing walk	-0.6388889	2.474473
Brisk walk	0.6388889	2.631162
Run	1.805556	2.974361

Table 2. Summary statistics of treatment groups vs response variables.

As shown in Figure 2 and Table 2, the boxplot and group means indicate a trend in change in overall negative mood across exercise intensities. Participants who engaged in high-intensity exercise (running) experienced the greatest improvement in overall negative mood with a mean change of 1.81 points. Moderate-intensity exercise (brisk walking) produced a lesser improvement of mean change in overall negative mood (0.64 points), while low-intensity exercise (relaxed walking) actually saw a worsening of overall negative mood, with a mean change of -0.64. These results would appear to indicate a preliminary dose-response relationship between exercise intensity and reduction of negative mood symptoms, with a significant effect beginning with a moderate intensity of exercise, and symptom relief increasing as intensity is increased.

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)
## factor(treatments)   2  107.6   53.81   7.420 0.000977 ***
## factor(block)        2   19.2    9.59   1.323 0.270939
## Residuals          103  747.1    7.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3. R output from ANOVA.

ANOVA gave an F-statistic of 7.42 and p-value of 0.000977 for the treatment effect. At a significance level of $\alpha = 0.05$, this indicates that at least one pair of treatment groups have statistically significantly different mean responses. In the context of this study, this means that at least one pair of exercise types/intensities have different effects on overall negative mood.

We also see that MS_{block} is only slightly larger than MS_{error} (9.59 vs. 7.25), which suggests that our response variable did not differ significantly across age brackets. Therefore, blocking on age was likely not necessary for this experiment.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = difference ~ factor(treatments) + factor(block), data = data)
##
## $`factor(treatments)`
##                                diff        lwr       upr     p adj
## relaxing walk-brisk walk -1.277778 -2.7873313 0.2317757 0.1141948
## run-brisk walk            1.166667 -0.3428868 2.6762202 0.1624543
## run-relaxing walk         2.444444  0.9348909 3.9539979 0.0005956
```

Figure 4. R output for Tukey's method for post-hoc testing.

Post-hoc testing using Tukey's HSD method revealed that participants who went on a 30 minute run had a mean change in overall negative mood 2.44 points greater than those who completed a 30 minute relaxing walk. This difference is statistically significant ($p = 0.0005956 < \alpha = 0.05$) and the confidence interval does not include zero, indicating that this effect is unlikely to be due to random chance. However, the difference in mean change in overall negative mood for participants who completed a run versus brisk walk, and relaxing walk versus brisk walk, were both not statistically significant ($p = 0.162$ and $p = 0.114$, respectively) and the confidence intervals both included zero. This indicates that while these groups did have different group means, the differences were not significant at $\alpha = 0.05$.

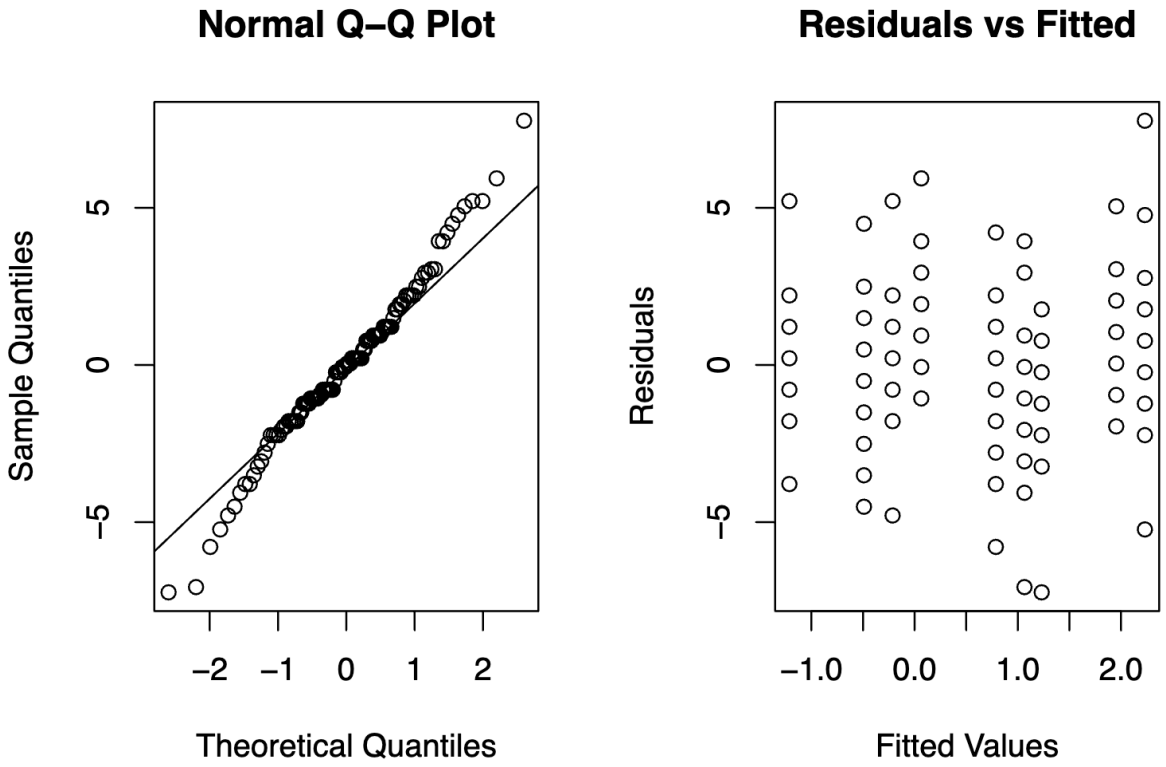

Figure 5. Residual plots from the model; all assumptions appear to be met.

Despite some slight deviation of residuals from the normal line, the Normal Q-Q Plot suggests the normality assumption is generally satisfied. The Residual vs Fitted Plot does not suggest any violation of model assumptions. The residuals do not appear to follow any pattern, and they are centered at zero.

Discussion

Since ANOVA found a p-value < 0.05 for a difference in treatments, we reject the null hypothesis that all three exercise types have the same effect on overall negative mood. Further analysis using Tukey's test showed that running and relaxed walking had significantly different effects on mood, but brisk walking was not significantly different from the other two exercise types. As such, we speculate that intensity of exercise might affect subjective ratings of overall negative mood, but only for low-intensity exercise compared to high-intensity exercise. Exercise of differing but more similar intensity levels, like low versus moderate and moderate versus high, does not appear to affect mood in significantly different ways. Since all three interventions were conducted outdoors and for 30 minutes each, the differing effects on mood due to treatment cannot be attributed to being outside or the duration of the exercise.

By comparing the Mean Squared value for the blocking groups and residuals, we conclude that there was likely little difference in change in overall negative mood between age groups. As such, it appears that all age groups could benefit equally from different intensities of exercise. This blocking variable was chosen to try and capture athletic ability; in the future, blocking by a more active indicator of physical fitness, such as time to run a mile, could result in a significant effect.

We used an RCBD method to investigate the effects of differing levels of exercise intensity, while blocking for age. While there did not seem to be differences by age, we did find significant differences in mood change between running and walking, while controlling for age. As such, exercise (particularly 30 minutes of running) could be an effective intervention to improve negative mood among the population of people born in Ironbard.

Limitations

The sampling methodology, while extremely robust in some ways, also produces some generalizability concerns. By using birth records from Ironbard, we were able to access the full population of individuals born on the island of Ironbard, so we had a true random sample from the population of interest. However, this population is not the same as the population of people living on Ironbard now. While we are able to generalize these results to the full population of people born on the island, we might not expect this group to be especially meaningful. Future studies with greater time and resources could benefit from sampling from people living on the island now, which might be more relevant.

These results are also not generalizable to the populations of the other islands, Providence and Bonne Santé. Since we only sampled from people born in Ironbard, we cannot use our results to make any claims about those born elsewhere. Future research that samples from all three islands could also investigate if any of the islands had different changes in overall negative mood due to exercise.

This study controlled for time spent exercising, as all interventions were 30 minutes in duration. However, this might be a key variable to investigate in future studies, as studies have found that less intense forms of exercise done for more time can be just as effective as more intense forms of exercise for less time. In particular, Choi et al. (2019) saw a significant decrease in the odds of becoming depressed for both 15 minutes of running, or one hour of walking. Perhaps, then, our results regarding intensity would have been moderated by length. At the same time, it is clear that the general direction of our results is supported by the literature, as the more substantial forms of exercise indeed improved mood overall mood.

Due to the way overall negative mood was measured, combining self-report responses for three distinct negative mood states into one overall metric, this study was unable to investigate if there were any differential effects on exercise on these individual mood states. Moreover, since we only considered negative moods, a more holistic approach could consider positive and negative mood states to see if there is a corresponding change in overall positive mood.

References

Choi, K. W., Chen, C.-Y., Stein, M. B., Klimentidis, Y. C., Wang, M.-J., Koenen, K. C., Smoller, J. W., & Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2019). Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. *JAMA Psychiatry*, *76*(4), 399–408. https://doi.org/10.1001/jamapsychiatry.2018.4175

Shacham, S. (1983). A Shortened Version of the Profile of Mood States. *Journal of Personality Assessment*, *47*(3), 305–306. https://doi.org/10.1207/s15327752jpa4703_14