Statistics 101B Final Project: Experimental Design to Explore the Islander Relationship between Music and Serotonin

Nurrachman Liu
UCLA
rachliu@ucla.edu

Justin Cox
UCLA
justincox@g.ucla.edu

Amaris Kim
UCLA
amarishkim@gmail.com

Avyay Kuchibotla
UCLA
avyayk@gmail.com

***Abstract*— We explore the design of an experiment to answer the question: does music effect discernible change in serotonin, in the Islanders of Edwardton city? We employ a split-plot repeated measures design. The main treatment factor is type of music a subject is exposed to, and the repeated measure of blood serotonin is repeated $3\times$. We block by gender and subject.**

We analyze using ANOVA for both a fixed-effects aov model as well as a random-effects lme model. Our results indicate that particular levels of music do indeed significantly change blood serotonin levels. Classical music and Sitting (control) are comparable and have the least effect, while Dance music has a statistically significant effect on blood serotonin levels. We found insufficient evidence for a significant effect from the other two music levels. The block-effect test showed strongly that our subject blocks were generalizable to the population of Edwardton.

I. INTRODUCTION

There is evidence that suggests music - particularly melodic music - can help mitigate the symptoms of a variety of physiological and neurological ailments. These ailments include Parkinson's Disease, arterial hypertension, dementia, and Attention-Deficit/Hyperactivity Disorder (ADHD). The benefits of music are likely related to the activation of brain monoaminergic circuits, which include the serotonergic pathways - the release of the "happy chemical" serotonin [1]. Serotonergic pathways likely help regulate the activity of dopamine, another neurotransmitter byproduct that is related to reward, emotion-driven behaviors, and motor control.

While observed clinical benefits typically occur in tandem to hearing melodic music, the concept of "melody" can be rather elusive. For instance, a Bruckner symphony (classical) can be atonal while a radio song can be highly melody-driven. Or vice versa, a Mozart symphony can be highly melodic while a pop song can be beat-driven. Some listeners may even find melody within atonality, much like an aural Rorschach test. To provide a more thorough overview on the specific effect of melody on neural pathways, we studied the effects of four disparate popular genres of music - classical, pop, dance, and country - on the release serotonin within subjects.

II. METHODS

A. Participants

For this experiment we employed the use of virtual participants from an online resource called the Island [2]. Each participant's consent was obtained first, with $\approx 20\%$ declining. Based on research from medical data journals, it is shown that Vitamin D causes the brain to release neurotransmitters such as dopamine and serotonin [3, 4]. With participants on the Island being exposed to differing levels of Vitamin D due to their natural environment, we made the decision to control for this factor and limited the selection of participants to be living in the city of Edwardton. Using GPower we determined that a sample size of 500 was needed.

To randomly select our subjects, we built an index of Edwardton city's population by crawling the pages using R. We found the total population size at the time to be ≈ 2000. We used R's random sample function to select 500 subjects out of this index of 2000 people.

B. Design

Below in Figure 1 and 2 is the factor and variable diagram outlining the Split-Plot Repeated Measures design for our experiment.

We decided to use the Split-Plot Repeated Measures design as we thought it would be important to see how the participants' blood serotonin levels changed with time as they were given the four treatments. Being limited by the Island, we were not able to have a perfect control for music listening. Therefore we felt that Sitting gave the best approximation of a control group for music listening. We also choose to block by gender, as research shows that there are natural differences in serotonin levels between men and women [5]. We also choose to vary the time factor by increments of 15 minutes as we felt that if the serotonin levels changed over time, we would be able to identify it over these time periods.

Response Variable	Blood Serotonin Levels														
Treatment (Music)	Sitting (Control)			Classical Music			Country Music			Dance Music			Heavy Metal Music		
Repeated Measures (Time in minutes)	1	15	30	1	15	30	1	15	30	1	15	30	1	15	30
Blocking (Gender)	Male						Female								

Fig. 2: Variable Diagram

C. Procedure

For our data collection process, we divided a random sample of 500 subjects from Edwardton city into five groups of 100. Each group was then assigned to one of the four experiment administrators, with one administrator handling two groups. The experiment administration entailed obtaining the consents, performing the main treatment, and subsequently measuring the responses.

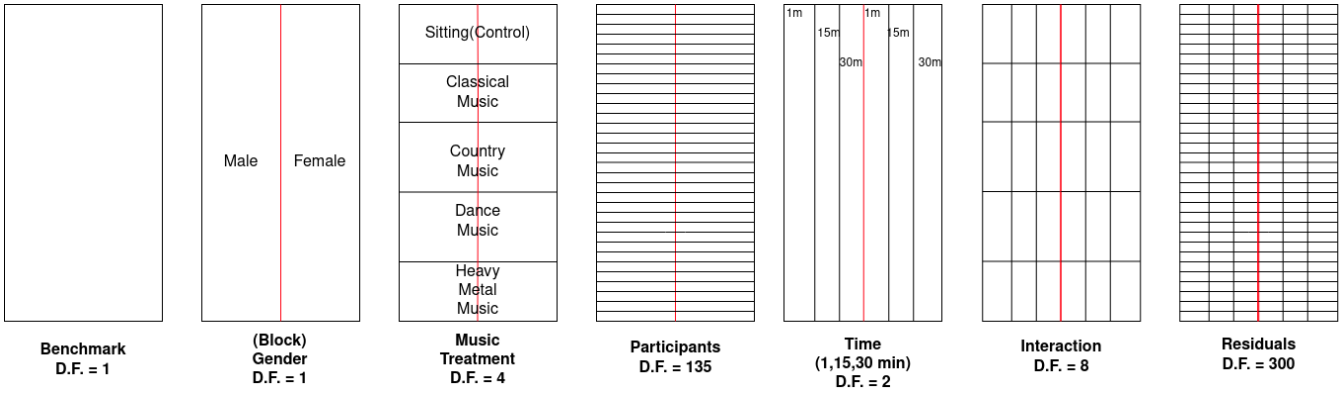

Fig. 1: Factor Diagram

Specifically, the following step-wise procedure was followed for administering the experiment:

1) Pre-treatment survey
2) Pre-treatment measurement of blood serotonin (response)
3) Main treatment application (Sitting, Music1, Music2, Music3, or Music4)
4) Post-treatment measurement 1 (1 minute elapsed since main treatment) of blood serotonin (response)
5) Post-treatment measurement 2 (15 minutes elapsed since main treatment) of blood serotonin (response)
6) Post-treatment measurement 3 (30 minutes elapsed since main treatment) of blood serotonin (response)

D. Instruments

Our subjects were the islanders of the Island. The surveys, main treatments, and blood serotonin responses were all applied as tasks offered by the Island. Our pre-treatment survey was configured to use all 61 questions available. The main treatment had 5 levels: Sitting, Music1 (Classical Music), Music2 (Country Music), Music3 (Dance Music), and Music4 (Heavy Metal Music). The response, blood serotonin (ng/mL), was collected using the task history results view for each islander. We then used R to analyze the collected data. Our second (inside) treatment factor was the three repeated-measures of the response. We note that although we took an initial response measure before main treatment application, this measure was not part of the repeated-measures factor and thus not part of the analysis.

III. Data Analysis

A. Statistical Analysis

We read our data files of our individual islander samples into R and merged them into one data frame. We fit a model using the aov() function to see if music, time, and the interaction between music and time have significant effects on the serotonin levels of our subjects. We blocked by gender and treated the island subjects as a nested factor. We checked if our model was valid by analyzing constant variance and normality through the graphs created with the plot() function. We also ran an interaction plot to see if there were indications of interaction between music and time. Then we built an ANOVA table to calculate the F-statistics and p-values to either reject or fail to reject our null hypothesis. To run a TukeyHSD post-hoc, we converted our model into a mixed effects model and see which factors are different from the others, if at all.

We used the following model for our split-plot repeated measures design:

$$aov(Blood.Serotonin \sim Main.Trt * Time.Measure + Gender + Error(Subject))$$

B. Sample Size Determination

We used the software G*Power to calculate the sample size needed for our repeated measures between factors design. We used a medium effect size of 0.25 and an alpha level of 0.05. In accordance with the minimum power of for most experiments, we had a power of 0.8, and the default correlation among repeated measures of 0.5 was kept. As our treatment factor has five levels of sit, classical music, country music, dance music, and heavy metal music, we had 5 groups. We measured each subject immediately after, 15 minutes after, and 30 minutes after treatment, so the number of measurements is 3. The minimum number of subjects needed is 135. We increased the sample size to 140 with 28 subjects in each treatment level to even out the number of islanders each member of our group had to sample.

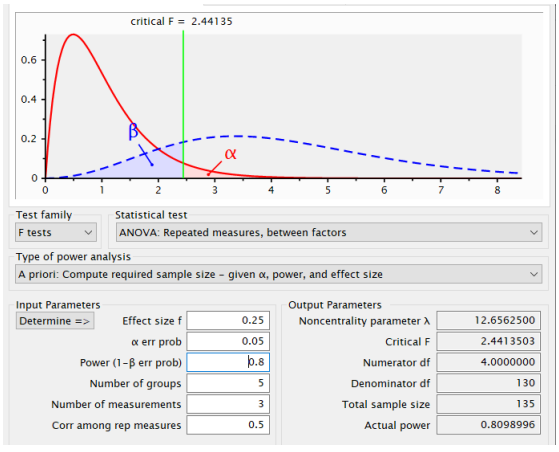

Fig. 3: Using G-Power to determine an approximate sample size for our design and desired power.

IV. Results

A. ANOVA Analysis

The ANOVA results are shown in Tables I and II below. For the ANOVA results between factors, we had a p-value of 0.0091 for the music treatment, which is much lower than our significance level of 0.05. This indicates that music does have a significant effect on serotonin levels for people living on the island, and that the kind of music has differing effects on serotonin levels.

For the ANOVA results within factors, we had p-values of 0.8419, 0.5682, 0.2390 for music treatment, time, and the interaction between music and time respectively. These p-values are well above our significance level, revealing that none of these factors are significant. Music does not have a significant effect within subjects, and time and their interaction do not either.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Main Treatment	4	24206.17	6051.54	3.51	0.0091
Gender	1	140088.11	140088.11	81.29	0.0000
Residuals	147	253324.08	1723.29		

TABLE I: ANOVA results, between.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Main Treatment	2	8.67	4.33	0.17	0.8419
Time	2	28.50	14.25	0.57	0.5682
Main Trt:Time	8	263.24	32.90	1.31	0.2390
Residuals	300	7548.92	25.16		

TABLE II: ANOVA results, within.

B. Residual Diagnostics

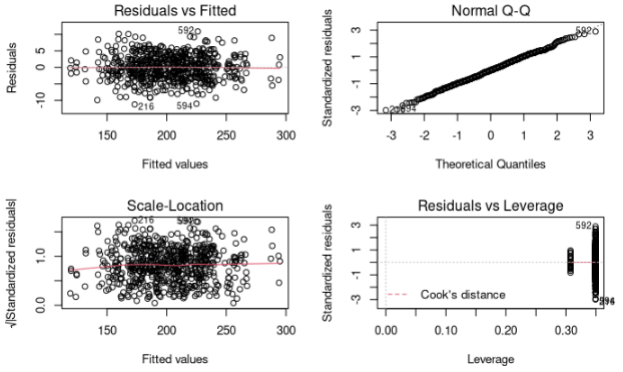

Fig. 4: Residual diagnostic plots for ANOVA. The assumptions of constant variance are maintained.

Residual diagnostics are shown in Figure 4. No discernible issues are seen; for example, no fan pattern in the residuals (top-left), and the normal fit (top-right) is very good up to 2.5 std-devs. Leverage plot (bottom-right) shows extremely few outliers and thus close to none bad-leverage points. Therefore, the assumptions of constant variance are maintained.

C. Interaction Plots

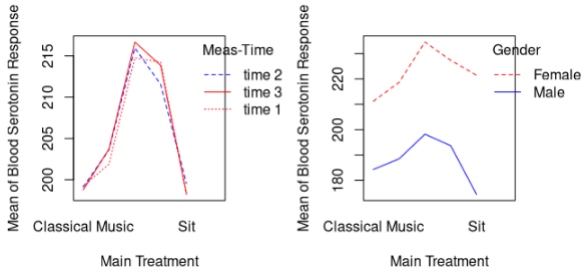

Fig. 5: Interactions between Main Treatment and Time (left), and Main Treatment and Gender (right). The p-value of either interaction terms is insignificant.

The main treatment does not significantly interact with the repeated-measure. It also does not interaction significantly with our blocking factor, gender, which is desired and confirmed the validity of using gender as our block.

D. Visualization

Figure 6 shows the response distributions. We note that that the ± 2 std-dev is large, indicating our significance result for the main treatment stems from only a few particular pairwise differences between means. We verify this below in the post-hoc analysis (Figure 8).

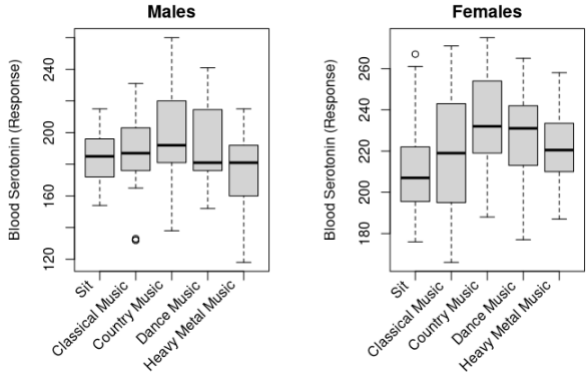

Fig. 6: Visualization of response data for males (left) and females (right).

E. Random Effects Model and REML

Since our Error structure is stratified, it is not possible to use the existing TukeyHSD post-hoc in R, which accepts only a regular AOV object, not a stratified AOV object (aovlist).

Therefore, one way to run post-hoc on a stratified model is to use a linear mixed-effects model. This made sense for our model, too, since the Subject factor was meant to be generalized. We therefore made the following fit, using the *lme4* package:

$$lmer(Blood.Serotonin \sim Main.Trt * Time.Measure + Gender + (1|Subject))$$

The LME results (Table III) match our fixed-effects model ANOVA results earlier. The results show that the control group (Sit, base) is similar in effect to Classical Music. Both of these, however, are significantly different than the other three types of Music treatments (t-values > 2). Gender is also strongly significant (t-value $= -9.1$), while the repeated-measures and most interactions are insignificant.

	Estimate	Std. Error	t value
(Intercept)	215.61	4.04	53.41
Classical Music	-0.49	3.57	-0.14
Country Music	7.38	5.07	1.46
Dance Music	17.33	5.64	3.07
Heavy Metal Music	10.34	5.07	2.04
time 2	1.35	1.27	1.06
time 3	0.13	1.27	0.10
Male	-35.19	3.87	-9.08
Classical Music:time 2	-1.45	1.80	-0.81
Country Music:time 2	0.39	1.80	0.21
Dance Music:time 2	-0.16	1.80	-0.09
Heavy Metal Music:time 2	-4.03	1.80	-2.24
Classical Music:time 3	-0.58	1.80	-0.32
Country Music:time 3	1.71	1.80	0.95
Dance Music:time 3	1.77	1.80	0.98
Heavy Metal Music:time 3	-0.52	1.80	-0.29

TABLE III: LME results for fixed-effects.

a) Diagnostics: Residual diagnositics for checking normality of residuals and constant variance assumptions for linear mixed effects model are in Figure 7. Assumptions are satisfied.

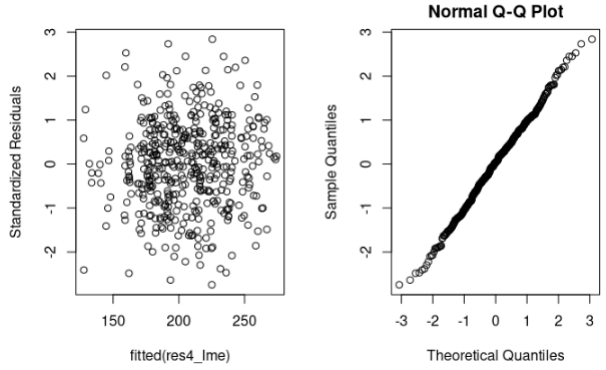

Fig. 7: Standardized residuals for checking homoscedasticity (left); QQ-Norm plot for checking residual normality (right).

b) ICC: From Table IV, the ICC is $\frac{563.70}{563.7+25.15} = 0.9573$, indicating that the repeated-measures within Subjects are strongly correlated, and that Subjects are very different from each other.

	Variance	Std. Dev.
Subject	563.70	23.742
Residual	25.15	5.015

TABLE IV: LME results for random-effects.

c) Block-Effect Test: The χ^2-test statistic was 2*($loglik_{\text{full}}$ - $loglik_{\text{red}}$) = 2*(-1701.432 - -2131.579) = 860.30, pval ≈ 0. We can conclude then, that the subjects are truly different from one another.

F. Post-hoc analysis

We ran Tukey's significance test on the lme model using the *lsmeans* package (Table V).

These results are consistent with earlier conclusions. Our control group (Sit) and Classical music are similar, and both are very different than the others. In particular, both of them

contrast	estimate	SE	df	t.ratio	p.value
Sit - Classical Music	1.1690	3.4276	447.71	0.341	0.9971
Sit - Country Music	-8.0771	4.9628	208.17	-1.628	0.4815
Sit - Dance Music	-17.8640	5.5486	169.12	-3.220	0.0132
Sit - Heavy Metal Music	-8.8230	4.9633	208.27	-1.778	0.3894
Classical Music - Country Music	-9.2461	4.9633	208.27	-1.863	0.3407
Classical Music - Dance Music	-19.0330	5.5498	169.21	-3.430	0.0067
Classical Music - Heavy Metal Music	-9.9920	4.9628	208.17	-2.013	0.2632
Country Music - Dance Music	-9.7869	5.5478	169.08	-1.764	0.3980
Country Music - Heavy Metal Music	-0.7459	3.4294	447.63	-0.218	0.9995
Dance Music - Heavy Metal Music	9.0410	5.5512	169.35	1.629	0.4813

TABLE V: Multiple comparisons results using Tukey.

are significantly different than Dance music. Country and Heavy metal music appear to be insignificant, as their 95% confidence interval spans 0.

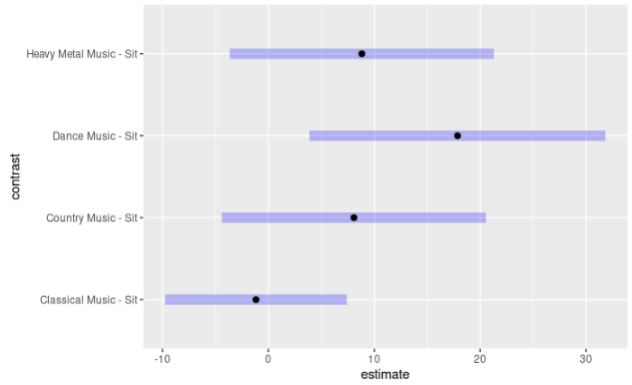

Fig. 8: LME effects differences (Tukey's) between treatment levels and control level (Sit) ($\alpha = 0.05$).

V. Discussion

A. Design Process and Iteration

After deciding on the SPRM design, G-Power reported 135 samples required for a power of 80%. We rounded up to 200, to margin for islanders declining to participate; a low-level blocking factor (gender); and disparities between cell counts (to ensure a balanced design).

Even with a large margin, we were limited by a minimum cell count of 21 subjects (Table VI).

	Classical Music	Country Music	Dance Music	Heavy Metal Music	Sit
Female	33	36	21	24	21
Male	39	36	51	48	51

TABLE VI: Cell counts for sample-size = 200. Only 21 usable Subjects per level (limited by females in Sit and Dance Music).

The ANOVA analysis showed that the main treatment was insignificant (Table VII). G-Power reported that our actual sample-size of $21 \times 5 = 105$ subjects resulted in only a $\approx$ 50% power. This implied a weak FTR of H_0.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Main Treatment	4	15393.46	3848.37	1.83	0.1274
Gender	1	65711.03	65711.03	31.29	0.0000
Residuals	114	239433.50	2100.29		

TABLE VII: ANOVA results, between, for a usable sample-size of 106 (raw count: 200).

We therefore decided to increase our samples to 500 Subjects total (100 per administrator). This yielded a cell count of $39 \times 5 = 195$ (Table VIII), meeting our power requirements.

	Classical Music	Country Music	Dance Music	Heavy Metal Music	Sit
Female	51	45	45	54	48
Male	42	48	48	39	45

TABLE VIII: Cell counts for sample-size = 500. 39 usable Subjects per level (limited by males in Heavy Metal Music).

B. Control group considerations

We wanted to select a task for a control group that would minimally affect blood serotonin response, but which would could also plausibly be given as a placebo to subjects. Our control group was an effective relative comparison point (Figure 8).

VI. Other Considerations

A. Blocking factors in addition to Subject

a) Gender: To reduce noise, we speculated that gender could be blocked. We found that while blocking by gender does not significantly improve the main treatment effect, it itself does have a significant effect on the blood serotonin response. Its interaction with the main treatment was found to be insignificant. These indicate that gender's impact on the MSE is small, despite it itself having a large mean squares. This indicated that we needed to pursue more aggressive options such as increasing sample size.

B. Noise Reductions

a) Replication by sampling same subjects across Islands: We investigated replication by each sampling the identical data set (sample-size=200, repl=4). However, this did not produce significant treatment differences (pval = 0.561). A likely reason is that replication is most helpful in extremely low sample-size designs such as latin squares, where it serves to boost the error degrees of freedom. In our case, we needed more *new* subjects to increase our treatment mean squares.

b) Operator block: We also investigated using our fully replicated data set (sample-size=200, repl=4) with an outside blocking factor for the operator (which person conducted the experiment). However, we found that operator blocking was insignificant (main treatment p-val = 0.670). A likely reason is that our operator error is small to begin with.

VII. Conclusions

In this project, we explored the design of an experiment to answer the question: does music effect discernible change in serotonin, in the Islanders of Edwardton city? We were able to conclude that Music does indeed effect a significant change in blood serotonin levels, but only for a few specific types of music. Classical music has the least effect, while Dance music has a statistically significant effect on blood serotonin levels. The results were generalizable to the population of Edwardton.

Acknowledgments

Thank you to Professor A. Almohalwas.

References

[1] MM Moraes, PCR Rabelo, Pinto VA, et al. "Auditory stimulation by exposure to melodic music increases dopamine and serotonin activities in rat forebrain areas linked to reward and motor control". In: *Neuroscience Letters* 673 (2018), pp. 73–78. DOI: 10.1016/j.neulet.2018.02.058.

[2] M. Bulmer et al. "Life on the Island: a Simulated Population to Support Student Projects in Statistics". In: *Technology Innovations in Statistics Education* 5.1 (2011).

[3] RP Patrick and BN Ames. "Vitamin D hormone regulates serotonin synthesis. Part 1: relevance for autism". In: *The FASEB Journal* (2014).

[4] Wayne A. Cass, Michael P. Smith, and Laura E. Peters. "Calcitriol Protects against the Dopamine- and Serotonin-Depleting Effects of Neurotoxic Doses of Methamphetamine". In: *Annals of the New York Academy* (2006). DOI: 10.1196/annals.1369.023.

[5] Rebecca Ashare, Catherine Norris, et al. "Individual differences in positivity offset and negativity bias: Gender-specific associations with two serotonin receptor genes". In: *Personality and Individual Differences* 55 (5 2013), pp. 469–463. DOI: 10.1016/j.paid.2013.04.009.