Botanical Garden Report

Qiwen Li #26020164

02/03/2020

```
data <- read.csv("~/Desktop/Stat 450 Project /AndreaByfuglien_data.csv",header = T, row.names = 1)
```

```
library(tidyverse)
```

```
## Warning: package 'readr' was built under R version 3.4.4

## Warning: package 'stringr' was built under R version 3.4.4

## Warning: package 'forcats' was built under R version 3.4.4
```

```
library(ggplot2)
library(broom)
```

Summary

Different variables have been related to pro-environmental behaviour, including valence (feeling positive and/or negative emotions), arousal (the state of being physiologically alert and attentive) and educational interventions.

An experiment was conducted at the UBC Botanical Garden to investigate how valence, arousal and education affect pro-environmental behaviour.

In this report, we try to explore roles of valence, arousal and education in pro-environmental behaviour. We use linear regression models to model the survey data and hypothesis tests to evaluate these effects.

As the previous proposal said.

Introduction

In this report, we will explore the relationship between arousal, education, valence and Z_Beh_Score (the standardized sum of response variables of donation, newsletters, volunteering and petitions). Arousal, education, valence as explanatory variables, and Z_Beh_Score as a response variable.

We use linear regression to fit the models, and we use hypothesis tests to determine if the variable has a significant linear relationship.

Moreover, we will remove the control group observations and then analyze walk and education separately; we will keep the control group observations and analyze the condition group; seeing if they give significantly different results.

To further explore, we will add age, gender, political views, and financial status as our explanatory variables, standardized donation, a standardized sum of newsletters and volunteering, standardized petitions, mean of ECO, mean of NEP, and mean of NR as our dependent variables, to fit full models and find the relationship between explanatory variables and dependent variables.

Data Description

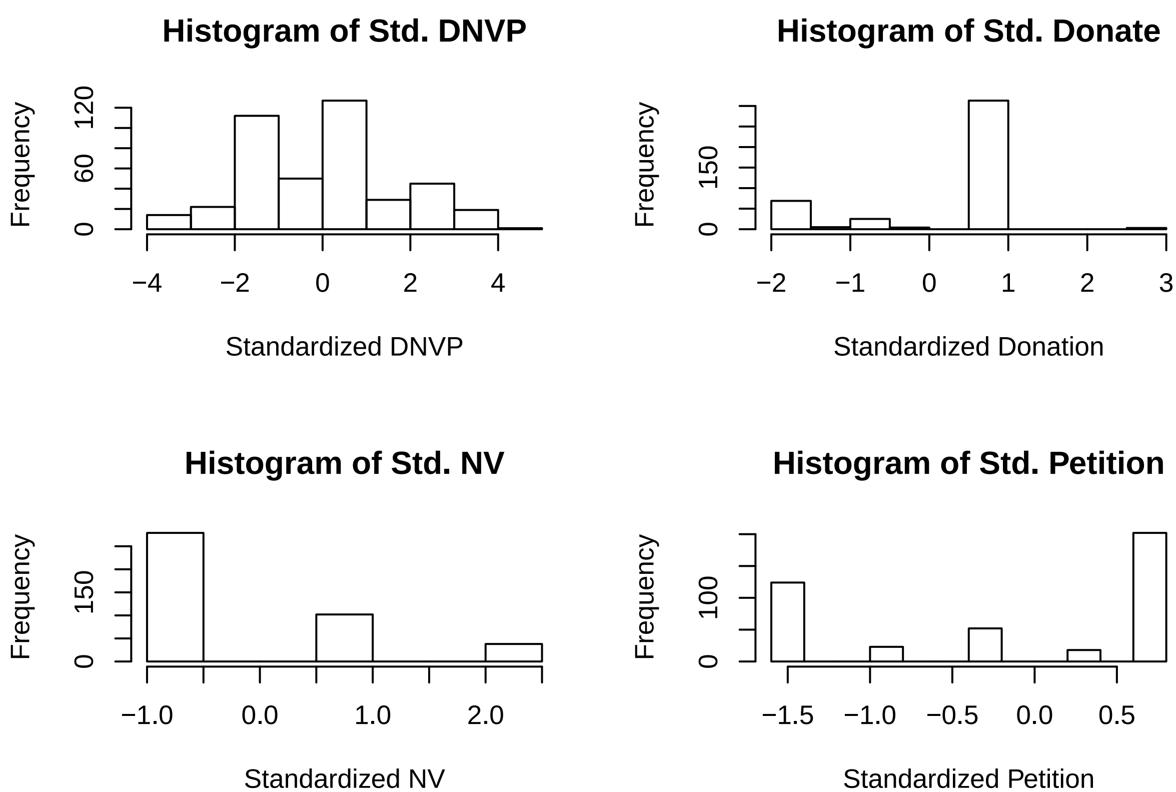
In this report, we will use walk, education, (or condition: the combination of walk and education) and valence as explanatory variables, and Z__Beh__Score as the dependent variable, which is the standardized sum of response variables of donation, newsletters, volunteering and petitions.

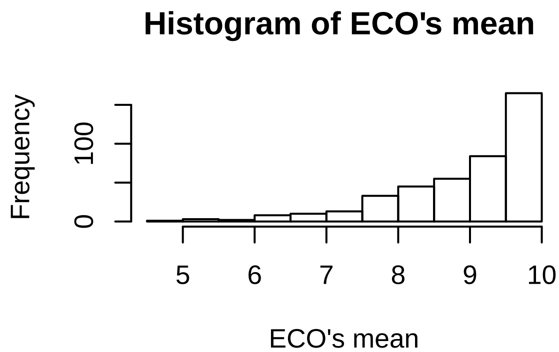
For the explanatory variables of walk, there are 85 participants in the control group, 169 participants in the ground walk group, 165 in the tree walk group.

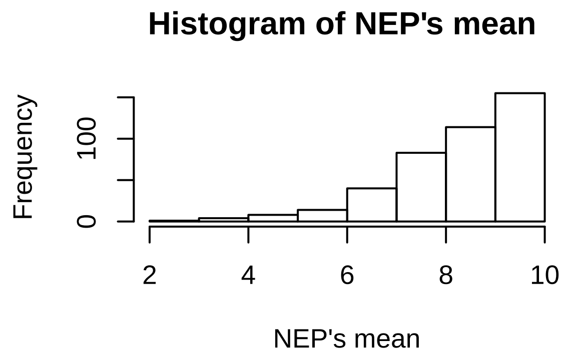
For the explanatory variables of education, there are 85 participants in the control group, 167 participants in the education group, 165 in the non-education group.

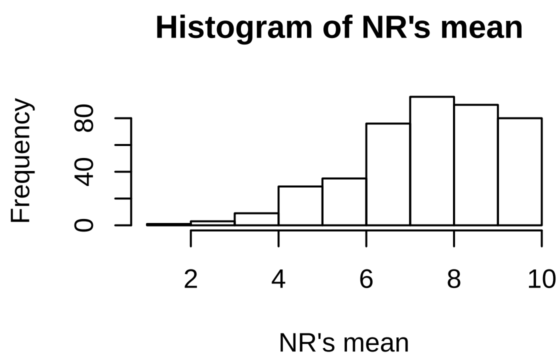
In the missing values, the number of age (missing 13 values), political views (missing 28 values), and financial status (missing 9 values) are large. Thus, we simply use the median of the rest data to substitute the missing value in the age, political views and financial status.

Also, the number of valence (missing 3 values), arousal (missing 3 values) and gender (missing 3 values) are not much; so we simply filter the participants corresponding to these missing values.

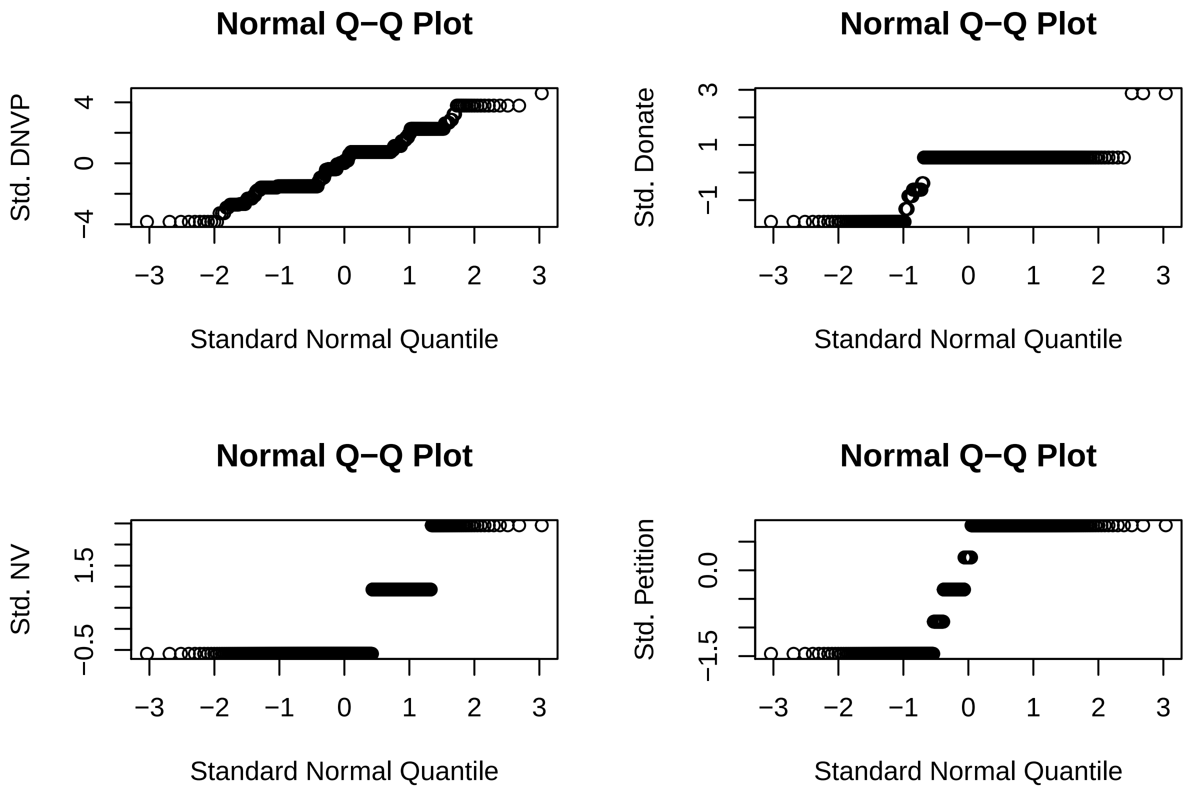


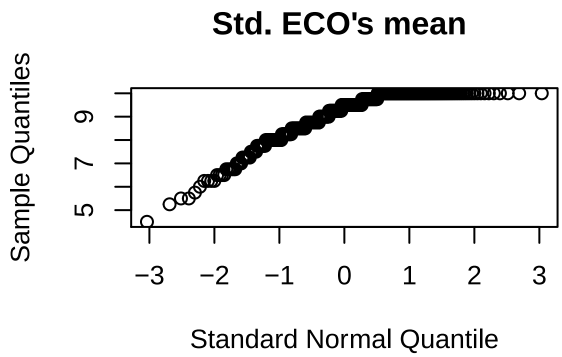


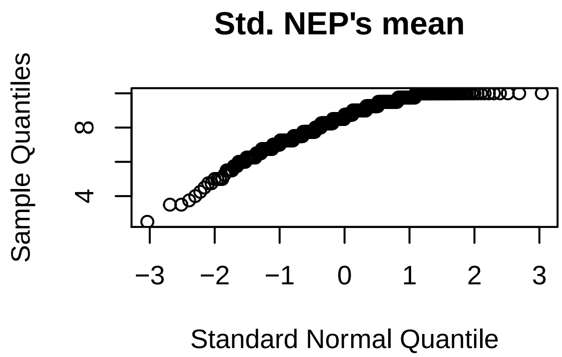


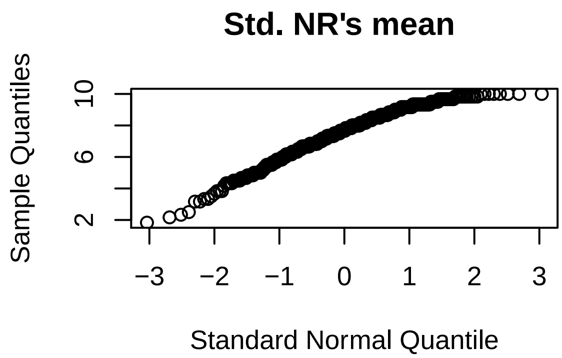


In the Figures above, we see that the histograms of the response variables (standardized sum of response variables of donation, newsletters, volunteering and petitions, standardized donation, standardized newsletters and volunteering, standardized petitions, mean of ECO, mean of NEP, and mean of NR) are not bell-shaped.

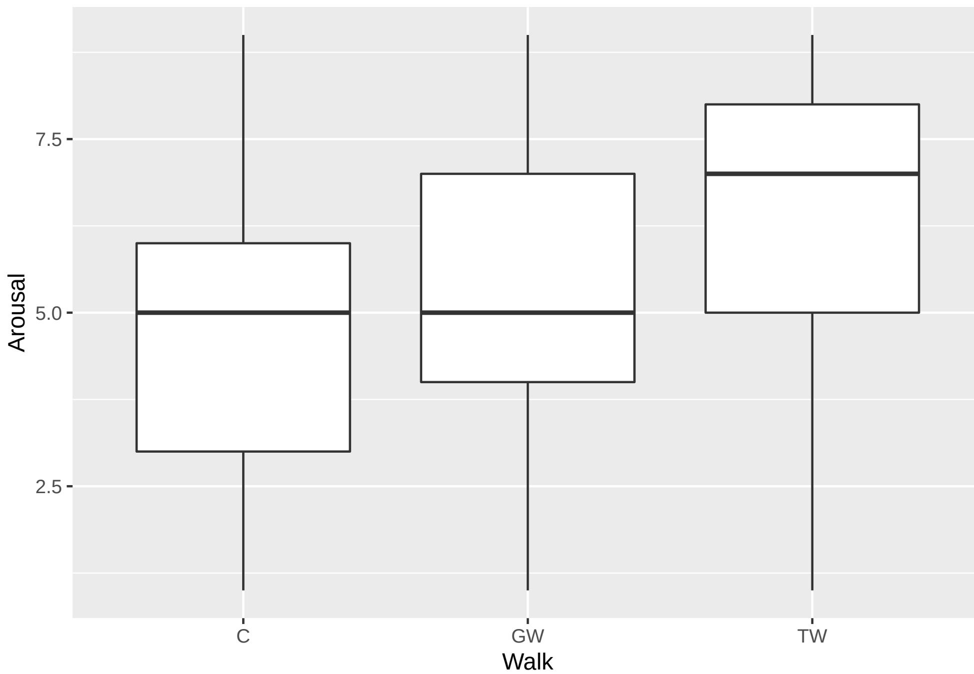

Normal Q−Q Plot
Std. DNVP
Standard Normal Quantile
Normal Q−Q Plot
Std. Donate
Standard Normal Quantile
Normal Q−Q Plot
Std. NV
Standard Normal Quantile
Normal Q−Q Plot
Std. Petition
Standard Normal Quantile







Furthermore, in Figures above, we use Q-Q plot to show that the dependent variables (standardized sum of response variables of donation, newsletters, volunteering and petitions, standardized donation, standardized newsletters and volunteering, standardized petitions, mean of ECO, mean of NEP, and mean of NR) are not normally distributed.



As shown in the above figure, different walks will affect the levels of arousal. For the control group, the level of arousal is the lowest; for the tree walk group, the level of arousal is the highest; for the ground walk group, the level of arousal is between the control group and tree walk group.

Methods

We use linear regression models to analyze the relationship between the dependent variable and explanatory variables.

For the dependent variables, we focus on the Z_Beh_Score (the standardized sum of response variables of donation, newsletters, volunteering, and petitions). We fit the linear regression model with explanatory variables: walk, education, and valence.

Meanwhile, we will remove control group observations and analyze walk and education separately; we will keep the control group observations and analyze the condition group; seeing if they give significantly different results.

To further our discussion, we add the mediator measures, including age, gender, political views, and financial status as our explanatory variables; and we add a standardized donation, a standardized sum of newsletters and volunteering, standardized petitions, mean of ECO, mean of NEP, and mean of NR as our dependent variables.

We simply use hypothesis tests to test our linear regression models, that is we test if the coefficients for each dependent variable are significant in the linear regression model. By doing so, we can evaluate the effect of the dependent variables, walk, education, and valence.

Results

We fitted four models for explanatory variables which are walk, education (or condition) and valence with the control group or without the control group.

1. For the linear regression model for three explanatory variables which are walk, education, and valence with the control group, valence is significant, the residual standard error is 1.731, and the adjusted R-squared is 0.01307. All the values of non-education are NA.
2. For the linear regression model for two explanatory variables which are condition and valence with the control group, valence is significant, the residual standard error is 1.733, and the adjusted R-squared is 0.01079.
3. For the linear regression model for three explanatory variables which are walk, education, and valence without the control group, valence is significant, the residual standard error is 1.728, and the adjusted R-squared is 0.002195.
4. For the linear regression model for two explanatory variables which are condition and valence without the control group, valence is significant, the residual standard error is 1.731, and the adjusted R-squared is -0.0006652.
5. For the full linear regression model for walk, education and the rest of explanatory variables with the control group, valence and political are significant, the residual standard error is 1.715, and the adjusted R-squared is 0.03047. All the values of non-education are NA.
6. For the full linear regression model for condition and the rest of explanatory variables with the control group, valence and political are significant, the residual standard error is 1.717, and the adjusted R-squared is 0.02807.
7. For the full linear regression model for walk, education and the rest of explanatory variables without the control group, political is significant, the residual standard error is 1.713, and the adjusted R-squared is 0.02013.
8. For the full linear regression model for condition and the rest of the explanatory variables without the control group, political is significant, the residual standard error is 1.715, and the adjusted R-squared is 0.01708.

We can find that in the full linear regression model, political is a significant explanatory variable for Z_Beh_Score.

Table 1: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-1.2172423	0.6575164	-1.8512729	0.0650345
as.factor(Walk)TW	0.0230911	0.1913663	0.1206643	0.9040312
as.factor(Education)NE	0.0947822	0.1908368	0.4966662	0.6197592
Valence	0.1484571	0.0785163	1.8907809	0.0595402

Table 2: ANOVA for Linear Regression Model

term	df	sumsq	meansq	statistic	p.value
as.factor(Walk)	1	0.0192928	0.0192928	0.0064591	0.9359936
as.factor(Education)	1	0.4252874	0.4252874	0.1423826	0.7061694
Valence	1	10.6784497	10.6784497	3.5750525	0.0595402
Residuals	326	973.7408332	2.9869351	NA	NA

Table 3: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-1.2172423	0.6575164	-1.8512729	0.0650345
as.factor(Walk)TW	0.0230911	0.1913663	0.1206643	0.9040312
as.factor(Education)NE	0.0947822	0.1908368	0.4966662	0.6197592
Valence	0.1484571	0.0785163	1.8907809	0.0595402

Table 4: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-0.9471583	0.7536062	-1.2568345	0.2097273
as.factor(Condition)2	0.0993787	0.2693256	0.3689908	0.7123779
as.factor(Condition)3	-0.0371960	0.2720420	-0.1367289	0.8913308
as.factor(Condition)4	0.0640833	0.2622807	0.2443311	0.8071305
Valence	0.1257692	0.0790588	1.5908309	0.1126320
Age	-0.0070778	0.0057503	-1.2308592	0.2192766
Gender	0.1134584	0.1787364	0.6347806	0.5260237
Political	-0.1482386	0.0639983	-2.3162899	0.0211719
Financial	0.0741401	0.0422406	1.7551844	0.0801817

Table 5: ANOVA for Linear Regression Model

term	df	sumsq	meansq	statistic	p.value
as.factor(Condition)	3	0.7056568	0.2352189	0.0799429	0.9708679
Valence	1	10.6212275	10.6212275	3.6097924	0.0583362
Age	1	3.4151892	3.4151892	1.1607062	0.2821281
Gender	1	2.0149906	2.0149906	0.6848265	0.4085445
Political	1	14.5520966	14.5520966	4.9457606	0.0268502
Financial	1	9.0643770	9.0643770	3.0806721	0.0801817
Residuals	321	944.4903256	2.9423375	NA	NA

Table 6: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-0.9471583	0.7536062	-1.2568345	0.2097273
as.factor(Condition)2	0.0993787	0.2693256	0.3689908	0.7123779
as.factor(Condition)3	-0.0371960	0.2720420	-0.1367289	0.8913308
as.factor(Condition)4	0.0640833	0.2622807	0.2443311	0.8071305
Valence	0.1257692	0.0790588	1.5908309	0.1126320
Age	-0.0070778	0.0057503	-1.2308592	0.2192766
Gender	0.1134584	0.1787364	0.6347806	0.5260237
Political	-0.1482386	0.0639983	-2.3162899	0.0211719
Financial	0.0741401	0.0422406	1.7551844	0.0801817

Table 7: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-0.9471583	0.7536062	-1.2568345	0.2097273
as.factor(Condition)2	0.0993787	0.2693256	0.3689908	0.7123779
as.factor(Condition)3	-0.0371960	0.2720420	-0.1367289	0.8913308
as.factor(Condition)4	0.0640833	0.2622807	0.2443311	0.8071305
Valence	0.1257692	0.0790588	1.5908309	0.1126320
Age	-0.0070778	0.0057503	-1.2308592	0.2192766
Gender	0.1134584	0.1787364	0.6347806	0.5260237
Political	-0.1482386	0.0639983	-2.3162899	0.0211719
Financial	0.0741401	0.0422406	1.7551844	0.0801817

Table 8: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-0.9471583	0.7536062	-1.2568345	0.2097273
as.factor(Condition)2	0.0993787	0.2693256	0.3689908	0.7123779
as.factor(Condition)3	-0.0371960	0.2720420	-0.1367289	0.8913308
as.factor(Condition)4	0.0640833	0.2622807	0.2443311	0.8071305
Valence	0.1257692	0.0790588	1.5908309	0.1126320
Age	-0.0070778	0.0057503	-1.2308592	0.2192766
Gender	0.1134584	0.1787364	0.6347806	0.5260237
Political	-0.1482386	0.0639983	-2.3162899	0.0211719
Financial	0.0741401	0.0422406	1.7551844	0.0801817

We have the linear regression model using a standardized sum of response variables of donation, newsletters, volunteering and petitions as our dependent variable, and condition, valence, age, gender, political, financial as our explanatory variables.

From Table 1-3

Table 1-2: the linear regression model for two explanatory variables which are condition and valence without the control group

Table 3: the linear regression model for three explanatory variables which are walk, education and valence without the control group

If we choose our significant level 0.05, the only p-values for valence are smaller than the significant level. Since we have the condition as a categorical variable, we can then perform the ANOVA test on this linear regression model. Based on the result of the ANOVA, the only valence shows a significant linear relationship to the standardized Beh_Score in our models. The estimated coefficient for valence is positive, which means that the valence has a positive linear relationship to the dependent variable.

From Table 4-8

Table 4-5: the linear regression model for three explanatory variables which are walk, education and valence with the control group;

Table 6: the linear regression model for two explanatory variables which are condition and valence with the control group

Table 7: the linear regression model for three explanatory variables which are walk, education and valence without the control group

Table 8: the linear regression model for two explanatory variables which are condition and valence without the control group

If we choose our significant level 0.05, the only p-values for political are smaller than the significant level.

Since we have the condition as a categorical variable, we can then perform the ANOVA test on this linear regression model. Based on the result of the ANOVA, only political shows a significant linear relationship to the standardized Beh_Score in our models. The estimated coefficient for political is negative, which means that the political has a negative linear relationship to the dependent variable.

Conclusions

In the reduced linear regression model with explanatory variables which are walk, education, and valence, valence is a significant explanatory variable for Z_Beh_Score. Valence shows a positive effect on motivating people to take action on climate change and sustainable development.

In the full linear regression model with explanatory variables which are walk, education and valence age, gender, political views and financial status, political is a significant explanatory variable for Z_Beh_Score. Political shows a positive effect on motivating people to take action on climate change and sustainable development.