Composting at Macalester

Laura Humes, Naomi Becker, and Emily Elkins

Intro to Statistical Modeling Final Project, Spring 2014

1. Introduction

Macalester is committed to sustainability, and is actively involved in reducing landfill waste. The Sustainability Office identified a goal of reducing unnecessary waste sent to landfills to zero by 2020. To this end, Macalester implemented a composting program in 2012. According to a waste assessment in 2010, approximately 45% of Macalester's waste output is compostable (Macalester Sustainability Office, 2010). Composting has significant environmental and economic benefits, allowing biodegradable waste to be turned into a useful byproduct rather than being contained in methane-producing landfills. Additionally, compostables are not subject to the 75% garbage tax in the state of Minnesota, creating an economic incentive to composting. Because composting has significant environmental and economic benefits, we decided to study the efficacy of Macalester's newly implemented composting program through a statistical study. We were particularly interested in determining *who* at Macalester is composting. In this paper, we present the relationships we found between demographic and composting behavior and hypothesize why certain demographic variables are significant to composting.

2. Data and Methods

We collected data from 105 Macalester students through an online survey conducted through Google Forms. We distributed the survey using our collective social networks, including word of mouth and Facebook postings. The survey results can be accessed through the R command *fetchGoogle("https://docs.google.com/spreadsheet/pub?key=0AqoMrZ2Z5D8sdFJBbmFjT1lLbF9XRlJZTlBMZVdFT0E&output=csv")*

Table 2.1: Variables Measured

Variable Name	Explanation
CompostFrequency	Measures how frequently the respondent composts on a scale from 1 to 5
RecycleFrequency	Measures how frequently respondent recycles on a scale from 1 to 5
PriorExposure	Measures level of prior exposure to composting before Macalester on a scale from 1 to 5
Confidence	Measures respondent's knowledge of what types of waste go into which bin (compost, recycle, trash) on a scale from 1 to 5
MoreEducated	Measures if respondent feels more educated about composting since coming to Macalester on a scale from 1 to 5
Important	Measures the importance of composting to respondent on a scale from 1 to 5
Gender	Gender of respondent (male, female, or write-in option)
Year	Respondent's expected year of graduation (2014 to 2017)
Major	Respondent's major area (Art/Humanities, Social Science, Natural Science/Math, or Interdisciplinary)
Region	Region of the United States respondent is from (East Coast, Midwest, West Coast, South/West, Outside of US)

Table 2.2: Summary of Variables

	CompostFrequency	RecycleFrequency	PriorExposure	Confidence	MoreEducated	Important	Gender	Year	Major	Region
Mean	3.771	4.276	3.371	3.79	3.048	3.838	Females: 75	2012=24	Arts/Humanities: 19	East: 27
Median	4	4	4	4	3	4	Males: 30	2015=12	Interdisciplinary: 9	Intern: 6
St. Dev.	1.171	0.882	1.527	1.133	1.259	1.17		2016=84	Natural science/Math: 37	Midwest: 39
Min	1	1	1	1	1	1		2017=27	Social science: 40	Southwest: 9
Max	5	5	5	5	5	5				West: 24

Though we attempted to collect data from a representative sample of 105 Macalester students, it was difficult to maintain an even distribution of students across all demographics we surveyed. For example, we observed an uneven distribution of major areas, likely due to advertising the survey within our own social networks, which tended to include more social science and natural science/math majors than other majors. Additionally, we had an uneven distribution of class years in our sample, again because within our own personal social networks, we know more underclass students than upperclass students. Despite these limitations, we believe that our data, while not entirely representative, spans a large enough sample that it can still be useful and relevant.

3. Results

In our initial proposal, we hypothesized that previous exposure to composting and region of the country that students are from would have significant effects on composting behavior. We therefore decided to study the interaction between previous exposure, area of the country, and composting behavior, in addition to other potentially useful predictors. However, performing a linear model analysis of the data in R revealed that other predictors were more important than these demographics in predicting composting behavior. We will detail these results below. To determined useful predictors, we first performed a linear model analysis of each predictor individually. These results are summarized in Table 3.1 To increase the strength of the model, we then examined the effect of multiple predictors. These results are summarized in Table 3.2.

Figure 3.1: Linear Model Analysis

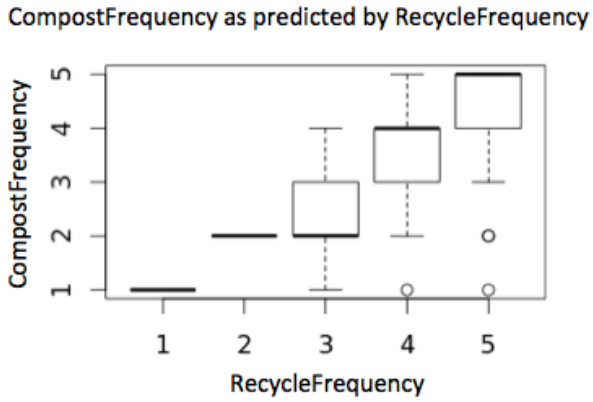
In order to see what the strongest relationships in our data were, we performed linear analyses of the predictors in our dataset with CompostFrequency, using the R command *summary(lm(CompostFrequency~Predictor, data=survey))*. Below is an example using RecycleFrequency as a predictor:

```
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.3591     0.4558   0.788    0.433
RecycleFrequency   0.7980     0.1044   7.643 1.15e-11 ***

Residual standard error: 0.9396 on 103 degrees of freedom
Multiple R-squared:  0.3619,        Adjusted R-squared:  0.3557
F-statistic: 58.41 on 1 and 103 DF,  p-value: 1.15e-11
```

Analyzing the linear model output for the relationship between CompostFrequency and RecycleFrequency reveals a very low intercept of 0.3591 and a relatively high coefficient of 0.798. This means that for every increasing degree (from 1 to 2, or 2 to 3, for example) that someone rates their recycling frequency, their composting frequency increases by 0.798. Therefore, there is a very strong relationship between recycling behavior and composting behavior. Indeed, the p-value of this model is

very significant, at 1.15E-11. We can see this same relationship playing out when we analyze the box plot of CompostFrequency as predicted by RecycleFrequency. To produce this plot, we used the R command *boxplot(CompostFrequency~RecycleFrequency, data=survey))*.



This boxplot confirms what we learned from the linear model analysis, that RecycleFrequency is a good predictor of CompostFrequency. As the mean of RecycleFrequency increases, so does the mean of CompostFrequency. There is very little variability in the lower end of the spectrum (in 1 and 2), those who do not recycle also do not compost. There is more variability in the middle part of the spectrum (3 and 4), but the relationship between RecycleFrequency and CompostFrequency continues to increase as we would expect. At the high end of the spectrum, those who rated RecycleFrequency a 5 generally also rate CompostFrequency as a 5, although there is some variablility and outliers. In general, though, this box plot confirms the results of the lm. We contrasted this the relationship between RecycleFrequency and CompostFrequency with other predictors using the same lm command. These results are summarized in the table below.

Table 3.1: Linear Model Analysis of Predictors without Interaction

Variable	Intercept	Coefficient	p-value	Adj. R-Squared	Residual Std. Error
RecycleFrequency	0.3591	0.798	1.15E-11	0.3595	0.9373
PriorExposure	2.88289	0.26355	0.000228	0.1146	1.102
Confidence	2.664	0.29216	0.002545	0.07543	1.126
MoreEducated	3.08439	0.22543	0.01038	0.05247	1.14
Important	1.23206	0.29682	8.26E-15	0.436	0.8795
Gender	4	0.8	0.00128	0.08746	1.118
Year	-672.88	0.3357	0.007142	0.05912	1.135
Major	3.89474	Inter: 3.89474	0.6131	-0.01152	1.177
		Sci/Math: -0.30014			
		SocialSci: -0.09474			
Region		INT: -0.9074	0.3608	0.00384	1.168
		MW : -0.4330			
		SW: -0.5185			
		WC: -0.1991			

In contrast to the relationship between RecycleFrequency and CompostFrequency, the relationship between PriorExposure and CompostFrequency, is much weaker, shown through the relatively high intercept (2.8829) and the very small coefficient (0.26355). This demonstrates that increases in prior exposure to composting have relatively little effect on actual composting behavior at Macalester. In fact,

one's recycling behavior has a much more significant effect on composting behavior than one's prior exposure.

Additionally, the variable Important has a strong relationship with CompostFrequency. The low intercept (close to zero) and high coefficient (0.436) means that frequency of composting behavior is directly positively correlated to how important one deems composting.

Figure 3.2 Linear Model Analysis and Analysis of Variance to Determine a Final Model

In order to determine the most useful combination of variables to use in a final model, we tried a variety of linear analyses using multiple predictors and compared these with and without using an interaction term (one predictor multiplied by another). We chose the predictors to use in the models by selecting a combination of the best predictors from Table 3.1. Below is a linear model and an analysis of variance output for our final model.

*summary(lm(CompostFrequency ~ RecycleFrequency + Important * Year, data = survey))*

```
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.499e+03  6.494e+02  -2.308    0.023 *
RecycleFrequency  4.765e-01  1.056e-01   4.514 1.74e-05 ***
Important         2.350e+02  1.609e+02   1.460    0.147
Year              7.436e-01  3.222e-01   2.308    0.023 *
Important:Year   -1.164e-01  7.984e-02  -1.458    0.148

Residual standard error: 0.7805 on 100 degrees of freedom
Multiple R-squared:  0.5725,        Adjusted R-squared:  0.5554
F-statistic: 33.48 on 4 and 100 DF,  p-value: < 2.2e-16
```

*summary(aov(CompostFrequency~RecycleFrequency+Important*Year, data = survey))*

```
                 Df Sum Sq Mean Sq F value   Pr(>F)
RecycleFrequency   1  51.57   51.57  84.652 5.59e-15 ***
Important          1  21.57   21.57  35.399 3.98e-08 ***
Year               1   7.16    7.16  11.751 0.000884 ***
Important:Year     1   1.29    1.29   2.125 0.148060
Residuals        100  60.92    0.61
```

Looking at the R-squared values from Table 3.1 and the results from Table 3.2, we chose the three most useful predictors and combined them them to create our final model. The R output for both the lm and the aov show that all three predictors (RecycleFrequency, Important, and Year) are useful to the model, with p-values that are significant (except for Important in the linear model, although that is approaching statistical significance). The p-value of the total model is incredibly significant, at <2.2E-16, and the Adjusted R-squared of the model is the highest of any model we tried, at 0.5554. Since the coefficient of year is positive, the higher the year the more likely the person is to recycle, meaning freshman recycle more than seniors. This can be explained both by the more recent placement of composting in the dorms (upper classmen who moved off campus this year never had the opportunity to live in a dorm with compost bins) and by the high likelihood of upper classmen to live off campus, and therefore have less access to composting. We also tried many other combinations of variables, summarized in the table below.

Table 3.2: Linear Model Analysis Using Multiple Predictors With and Without Interaction

Variable 1	Variable 2	Intercept	Coefficient (Variable1)	Coefficient (Variable 2)	Coefficient (Interaction)	p-value (model)	p-value (variable 1)	p-value (variable 2)	p-value (interaction)	Adj. R-Squared	Residual Std. Error
RecycleFrequency	Important	0.0601	0.4442	0.4721	N/A	<2e-16	0.00012	0.00000016	N/A	0.504	0.825
RecycleFrequency	Important	0.9101	0.2334	0.1706	0.0711	1E-15	0.32	0.57	0.3	0.158	0.824
RecycleFrequency	PriorExposure	0.3387	0.7368	0.0836	N/A	5.11E-11	4.5E-09	0.21	N/A	0.359	0.937
RecycleFrequency	PriorExposure	0.36088	0.73138	0.07436	0.00216	3.23E-10	0.00093	0.8124	0.97594	0.353	0.942
RecycleFrequency	Year	-704.7976	0.8045	0.3498	N/A	2.09E-13	9.5E-13	0.0004	N/A	0.425	0.888
RecycleFrequency	Year	168.4521	-202.6433	-0.0834	0.1009	1.02E-12	0.38	0.87	0.38	0.424	0.889
PriorExposure	Important	1.2475	-0.0265	0.6809	N/A	1.76E-13	0.7	1.4E-11	N/A	0.427	0.438
PriorExposure	Important	1.29918	-0.04694	0.66648	0.00521	1.27E-12	0.84855	0.00062	0.93101	0.421	0.891
Year	Important	-516.96261	0.25709	0.64257	N/A	5E-15	0.0065	2E-14	N/A	0.4663	0.8552
Year	Important	-1418	0.7042	232.2	-0.1149	1.5E-14	0.0479	0.1893	0.1905	0.4701	0.8521

4. Conclusion

Based on our calculations, we concluded that people who recycle are more likely to compost. Contrary to our initial hypothesis, prior exposure to composting has a negligible effect on one's likelihood of composting. We were also surprised to find that education since coming to campus about composting does not have a significant effect on one's likelihood to compost. Similarly, one's place of origin does not affect one's likelihood of composting, nor does prior experience. This is encouraging news for the Zero Waste Program—despite the great diversity in backgrounds of Macalester students, place of origin or prior experience to composting does not significantly impact composting behavior. Instead, one's view of how important composting is directly affects the likelihood of the person to compost. In fact, it is the most important predictor we found when analyzing our data Therefore, we encourage the Macalester Sustainability Office to emphasize the importance of composting in their Zero Waste Program, in order to allow students to see that composting has important environmental and economic benefits, therefore increasing the personal relevance of composting to the individual as well. In conclusion, basic education about composting is not enough; Macalester needs to instill the importance of composting in the student body in order to create a composting culture on campus.

5. References

Macalester Sustainability Office (2010). *Sustainability at Macalester: Composting*. Retrieved from: www.macalester.edu/admdept/sustainability/public.www/composting/index.html