

### **Student Categorical Proposal (Step 1):**

For the final project in this class, I plan to write about some analyses that I recently conducted using methods from this course. The purpose of the project was to investigate the extent to which class and gender differences exist in self-provisioning behavior, labor in which the laborer is simultaneously producer and consumer. In particular, I tested the hypothesis that higher levels of social status, operationalized in terms of income and business ownership separately, predict less time spent self-provisioning in a variety of domains. As a secondary hypothesis, I proposed that the well-established gender differences in self-provisioning (e.g. Cast & Bird, 2007) would be consistent across social strata.

The data for this project came from the American Time Use Survey (ATUS) and were collected by the Department of Labor from 2003-2008 as a way of studying how Americans spend their work and leisure time. I found support for both of my hypotheses across a number of domains using a series of negative binomial regressions and testing models testing these three predictors (i.e. wage, business ownership, gender) and their interactions. For the purposes of this project, I am also considering looking at certain domains in terms of either the presence or absence of the activity (e.g. mowing the lawn) and looking at these using the same predictors but in terms of logistic regression.

### **Professor Comments on Proposal:**

This sounds like it is on its way to being an excellent project. I can't completely envision all your analyses, but it sounds like you are thinking about a whole series of analyses. As it sounds like you are working towards a publication, the series of analyses certainly will be sufficient for this class. I encourage you to consider using this class project as a draft of your submission, rather than just an step in-between the analyses and the final paper that will be submitted. As you are using the negative binomial, please be sure to check that the dispersion parameter warrants used of the negative binomial rather than a Poisson distribution.

### **Retrospective Reflection:**

This proposal was somewhat incomplete but suggested that an appropriate data set had been selected for the question being addressed. Because of the size of the data set, it was clear that the student would have a lot of flexibility in analysis, and if anything the challenge would be in selecting a clear and concise set of analyses. Because there were relatively few details about the planned analysis, it was difficult to know for sure that the paper was on the right track, but the comments were intended to get the student thinking about their analysis (negative binomial comments).

## Student Paper Draft Excerpts (Step 2):

[...]

Americans also engage in *self-provisioning*: labor for which the laborer is both producer and consumer. People grow their own food in the garden, cook their own meals, do their own laundry, etc. In short, there are three labor markets into which Americans invest time: 1) Formal, or on-the-books, labor, 2) Informal, or off-the-books, labor, and 3) Self-provisioning.

[...]

Specifically, there are two primary research questions of interest. First, what is the relationship between gender and class? Do the effects of gender remain consistent across social classes in all domains? If not, of course, we will need to make sense of those differences. Secondly, what is the relationship between social class and self-provisioning? We have reason to believe that the relationship between class and self-provisioning is somewhat complicated: sometimes we expect wealth to reduce the time spent self-provisioning, but then there are forms of self-provisioning that require immense wealth to do.

[...]

Methods

[...]

Results

[...]

### *Cleaning, Laundry, and Sewing*

For this factor, and for all others, we have a DV that is a count variable (minutes spent doing the task). We see that for this factor, however, a Poisson distribution is not appropriate ( $M = 9.01$ ,  $SD = 19.15$ ). In fact, the AIC for the full Poisson model is infinite, indicating that this is a poor choice. For this factor, we used Negative Binomial regression, which allows us to account for the overdispersion in the data<sup>1</sup>. Predicting minutes spent on this factor by gender, business, ownership, and wage, we see a significant two way interaction we do not see a three way interaction, but two significant two way interactions which qualify our main effects.

First, there is a significant interaction between gender and business ownership,  $\beta = -.375$ ,  $z = 4.23$ ,  $p < .0001$ . In particular, we see that for business owners, women are doing more ( $\beta = 1.45$ ,  $z = 18.57$ ,  $p < .0001$ ) and that this effect is in the same direction for non-business owners ( $\beta = 1.08$ ,  $z = 36.92$ ,  $p < .0001$ ). The significant interaction suggests that the gender gap is significantly smaller for non-owners.

There is also a significant gender by income interaction ( $\beta = -1.23 \times 10^{-6}$ ,  $z = 2.56$ ,  $p < .01$ ). In particular, we see that for men, income is not a significant predictor of the amount of time spent on cleaning, laundry, and sewing,  $z = .687$ ,  $p > .49$ . For women, however, we observe the predicted effect of wage,  $\beta = -8.33 \times 10^{-7}$ ,  $z = 2.84$ ,  $p < .01$ . For women, higher weekly income predicts less time spent on cleaning, laundry, and sewing.

[...]

Discussion

[...]

As a way of approaching these findings, I would like to propose a few clusters that we should consider for the sake of clarifying these findings.

### *Traditionally Feminine Self-Provisioning*

Factors 1 and 2 have parallel results in the regression analyses above. These factors include cleaning, sewing, laundry, cooking, and kitchen tasks. In short, these are highly gendered tasks that are typically considered the domain of women's work. We see, correspondingly, that women are doing much more of this work compared to men.

Interestingly, the gender gap is somewhat narrower for lower class (i.e. non-business owning) respondents. This fits sociological analyses (Nelson & Smith, 1999) which suggest that in some ways, gender roles are a luxury. If, for example, both partners in a household are working, then who does the cooking or laundry may depend more on who has time and less on gender roles.

---

<sup>1</sup> All of the analyses are conducted using Negative Binomial regression through the Zelig package (Imai, K., King, G. and Lau, O. (2007). "negbin: Negative Binomial Regression for Event Count Dependent Variables" in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," <http://gking.harvard.edu/zelig>

[...]

#### *Traditionally Masculine Self-Provisioning*

We also see that a few factors stand out as more common among men. In particular, interior maintenance, vehicle maintenance, and appliance maintenance all have a strong main effect for the gender of the participant. Fixing up around the house and working on the car fall under traditionally masculine self-provisioning.

With regard to interior maintenance, we observe a main effect of wage such that those who make more money spend more time engaged in interior maintenance. This is perhaps due to the fact that wealthier individuals are more likely, for example, to own a house and to be concerned with maintenance. Wealthier individuals are presumably also more likely to own things that are worth maintaining, whereas poorer individuals might rely on used goods or more disposable products.

Wage, however, has the opposite effect when it comes to vehicle repair and maintenance. For men, we see that making more money means spending less time in the garage. Wealthier men may prefer to hire out labor for auto repair by taking broken vehicles to a mechanic or dealer, while others may try to save money by doing this work for themselves. The effect may also be the result of the kinds of cars that people are able to afford. Wealthier individuals are going to have the money to buy newer cars that require less maintenance (and still fall under warranty) whereas lower class individuals may depend more on older used cars.

[...]

#### *High Investment Self-Provisioning*

Some forms of self-provisioning are only necessary because the people who engage in them can meet the initial investments. For example, note that lawn and garden care, pet care, and exterior maintenance all require serious investment. To be able to garden, for example, one needs sufficient land and time. Likewise, it takes considerable disposable income to afford one pet, let alone more than one.

As a result, we see that in general, for both men and women, owners and non-owners alike, increased income predicts more time spent in high-investment self-provisioning. Households that have more money potentially have more space to garden, more pets in the house, bigger yards to mow, etc.

[...]

#### *Conclusions*

[...]

#### **Professor Comments on Proposal:**

[...]

My primary concern is that while the poisson distribution or negative binomial distribution does seem to be appropriate given the interval nature of the data, it seems likely that there is going to be a larger preponderance of zeros in the data. While a larger percentage of people may mow their lawn, on any given day it is unlikely that a person has mowed their lawn. I suspect that the data consist of a large number of zeros, followed by a distribution of remaining scores. It would be interesting to see whether the pattern of results changes if one uses a method designed to model the zeros differently from the other values, before fitting the models to the non-zero values. A table indicating the percentage of people who gave non-zero responses would also be helpful --- as this might call into question how much can be inferred from these data. If only 20 people reported vehicle repair and maintenance, we might not want to put much credit in the interpretation of the results presented.

Minor concerns & suggestions:

-The author should address the practical significance of the results. Given the large sample size, it is not surprising that significant p-values are very small --- but in practical terms, how large are these effects? By how many minutes do genders differ on these tasks? If we extrapolate out to a week, how many more hours a week are women spending cleaning (for example)?

-Along the same lines as the previous comment, it would also be helpful to rescale predictors. For example, a 1 unit change in weekly income is difficult to understand. It does a poor job of conveying the effect. What are the effects for something like a \$100 or \$500 change in weekly income?

[...]

-Does the survey include information about the number of hours individuals work outside of the house? If so, the author needs to control for this information --- someone who doesn't work outside of the house will probably do much of the self-provisioning in the household. Furthermore, households with only one income are likely to do much more self-provisioning to make ends meet.

-The effect of income seems likely to be nonlinear. Please test the effect of a linear and quadratic effect of income. Please include p-values indicating not only the effects of the linear and quadratic components individually, but tested simultaneously. That is, test the significance of including both effects compared to a model without either effect, so as to get an impression as to whether income as a whole is a useful predictor.

[...]

### **Retrospective Reflection:**

This paper did an extensive number of analyses, as eight different areas of self-provisioning behavior were examined. While the student put a lot of time into completing each set of analyses for each of the eight sections and writing it up, unfortunately there was a key issue with the analysis. The presence of zeros in his data, bound to occur because people don't mow their lawn every day, called into question the relationships that had been observed. While the paper was also very complete in addressing statistical significance, it lacked consideration of the practical meaning of the results. The paper was reasonable at this stage, but would move into being excellent if these key issues were addressed.

### Student Final Paper Excerpts --- Key Changes (Step 3):

[...]

#### Results

Prior to analyzing time use spent self-provisioning, analyses should be run to verify the predicted differences between gender and business ownership groups on average weekly wage. Because the distribution of income is best fit by a Gamma distribution (Salem & Mount, 1974; McDonald & Jensen, 1979), a GLM function coding for business ownership, gender, and their interaction should be modeled...in a better version of this paper.

Turning to the analyses of self-provisioning, we see a number of important findings using our three predictors. I will discuss each factor separately below. The purpose of these analyses is specifically to test the unique effects of gender and social class, and hence time spent at work will be controlled throughout as a covariate. To enhance the interpretability of the results, I have also recoded wage by dividing by 100, thus all effects of wage will be reported in units of 100 dollar per week increments. *Cleaning, Laundry, and Sewing*

For this factor, and for all others, we have a DV that is a count variable (minutes spent doing the task). We see that for this factor, however, a Poisson distribution is not appropriate because the mean should approximate the variance, but this is clearly not the case ( $M = 36.05$ ,  $SD = 76.61$ ). As a result of the overdispersion in these data, I will therefore use negative binomial regression which can account for this additional variability.

It was also noteworthy that 64.51% of participants (31432 out of 48724) reported a total of 0 minutes on cleaning, laundry, and sewing in the previous day. Given that less than half of the sample has non-zero value for this count, it was appropriate to use zero-inflated negative binomial regression, which models specifically for those participants who have a count on this variable<sup>2</sup>.

Specific models were compared using a series of likelihood ratio tests. It was found that the optimal model includes all main effects for gender, wage, and business ownership as well as a gender x business owner interaction (and hours at work as a covariate). This model was significantly better than the intercept only model ( $\chi^2(10) = 9542$ ,  $p < .0001$ ) and reduced models including only gender ( $\chi^2(6) = 137.76$ ,  $p < .0001$ ), wage ( $\chi^2(6) = 4462.31$ ,  $p < .0001$ ), ownership ( $\chi^2(6) = 4633.70$ ,  $p < .0001$ ), or the model containing all main effects ( $\chi^2(2) = 25.49$ ,  $p < .0001$ ). It was also found that no more complex models significantly differed from this model (all  $ps > .05$ ), indicating that this is the most parsimonious model.

Once this was determined, a series of models were estimated which included the main effect for a quadratic effect of income, as well as possible interactions between the quadratic wage term, gender, and business ownership. The optimal model in this case added a main effect for the quadratic term as well as an interaction between gender and the quadratic term to the previous model. This model was significantly better than the model without quadratic terms for wage ( $\chi^2(4) = 96.23$ ,  $p < .0001$ ), the model that included only the quadratic term for wage ( $\chi^2(2) = 17.69$ ,  $p < .001$ ). No additional interaction terms significantly improved upon this model (all  $ps > .05$ ). Hence, the optimal model is as follows: time = wage + ownership + gender + (gender \* ownership) + wage<sup>2</sup> + (wage<sup>2</sup> \* gender). This zero-inflated model was tested against its non-zero-inflated negative binomial equivalent<sup>3</sup>, and found to be a significantly better fit to the data (Vuong test statistic for non-nested models = 92.99,  $p = 0$ ).

It is worth noting that not all of the effects in this optimal model are actually significant (see Table 1 for a summary). Notably, the gender x ownership interaction ( $z = 1.95$ ,  $p = .051$ ) is only marginally significant, and the interaction between gender and the quadratic term for wage is also non-significant ( $p = .86$ ). There are also no significant effects for the quadratic wage term ( $p = .13$ ) nor for business ownership ( $p = .11$ ).

The only effects worth noting in the final model are a main effect for gender and a main effect for wage. In particular we see that wage (in \$100 per week increments) is a significant predictor,  $\beta = -.013 \pm .0067$ ,  $z = 3.93$ ,  $p < .0001$ . To put this into perspective, someone making 100 dollars more than someone

---

<sup>2</sup> All zero-inflated-models were run using the `zeroinfl` command as part of the `pscl` package on R (Jackman, Tahk, Zeileis, Maimone, & Fearon, 2010).

<sup>3</sup> All non-zero-inflated models were run using Imai, K., King, G. and Lau, O. (2007). "negbin: Negative Binomial Regression for Event Count Dependent Variables" in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," <http://gking.harvard.edu/zelig>

else will spend .98 ( $e^{-.013}$ ) times fewer minutes cleaning, doing laundry, and sewing. At a gap of 500 dollars per week, this is a difference of .93 times, or the difference of slightly more than four minutes per hour ( $60 * .93 = 55.8$ ).

There is also a significant effect for gender,  $\beta = .2388 \pm .097$ ,  $z = 4.84$ ,  $p < .0001$ . As we might expect, women are spending 1.27 times as many minutes cleaning, doing laundry, and sewing. In terms of minutes, this means that a woman would be expected to spend 76.2 minutes ( $1.27 * 60$ ) working at this task for every hour that a man spends.

[...]

## Discussion

[...]

### *Capital Required or Class Advantage or Quadratic Effects?*

With regard to social class, I suggested at least two ways we could think about the relationship between social class and self-provisioning labor. Many forms of self-provisioning labor require investments of capital to even make the labor possible, as in the case of gardening, which requires both time and land. However, for other forms of self-provisioning, such as cooking, it appears that wealthier individuals should be expected to do less as they can afford, e.g. to go to restaurants more often. Poorer individuals should be expected to self-provision because this offers one concrete way of saving money (by sacrificing time instead). It is cheaper to make a box of macaroni and cheese than to order pizza, but the former takes more time than the latter.

Given these two competing approaches to thinking about the relationship between social class and self-provisioning, it seems reasonable to expect that perhaps the effect of wage is quadratic. Perhaps as people make more money, they have more access to domains of self-provisioning (e.g. pools to clean), but that after a point, increased wealth leads to decreased self-provisioning (e.g. as people hire someone to clean their pool). These analyses attempted to test all three of these potential trends by looking for simple directional effects of wage and business ownership, as well as quadratic effects of wage within these data.

[...]

### *Are the Gender Gaps Stable Across Classes?*

Finally I would like to return to the second research question behind this project: are gender differences in self-provisioning stable across social classes? In general, we see that the answer is yes: for a number of factors, there are main effects of gender, but not significant interactions with either business ownership or income.

In particular, we see main effects of gender on a number of factors that are traditionally gendered. Men spend more time engaged in vehicle repair and lawn and garden work, and as we might expect, women spend more time cleaning and cooking. Again, it is worth pointing out that cleaning and cooking behaviors are far more common of course, and that the time use disparities here are by no means equal over the long run.

There are only two factors for which notable gender x class interactions arise. First, with regard to cooking, we see that the gender gap is smaller between business owning and non-business owning households. In both cases, women are doing more of this essential, everyday labor, but the gap shrinks for lower class families. This fits with analyses of self-provisioning practice that have demonstrated the extent to which gender gaps are in some ways a luxury for upper class families (Nelson & Smith, 1999). Poorer households may have, for example, both partners of a heterosexual couple working, which may require men to take up more of the cooking. Upper class households may have more freedom to be single-earner households, and it is normative for the man in the household to be that single-earner. This leaves cooking to the non-earning partner: gender normatively this is the woman.

Secondly, there is an interaction between gender and ownership with regard to exterior maintenance (e.g. installing windows, cleaning the pool). In upper class households, there is no gender gap for this form of labor, while we do see a gap for lower class households (in which men are doing more of this sort of "handyman" work). I attempted some post hoc analyses to suggest that the absence of a gender gap among upper class families is the result of the fact that upper class individuals regardless of gender are simply doing less exterior maintenance. The gender gap here is that men are doing more, but specifically lower class men. Again, we might expect that lower class men are motivated to perform their own home repair and maintenance as a way to save money in a way that upper class men are not.

### *Conclusions*

[...]

**Retrospective Reflection:**

In the revised paper, the student did an extensive revision of the analysis, worked to consider the practical significance of the results, and rewrote the discussion of the result to be clearer and more in line with his questions of interest. These extensive improvements led the paper to meet most of the requirements for an excellent paper (see rubric). The student has since worked to revise the paper further for publication.