

## Real Change Data Analysis

Camille Zaug and Salvatore Calatola-Young

June 3, 2020

**Guiding Question:** Is there a relationship between a vendor's housing situation and their hours spent at their post?

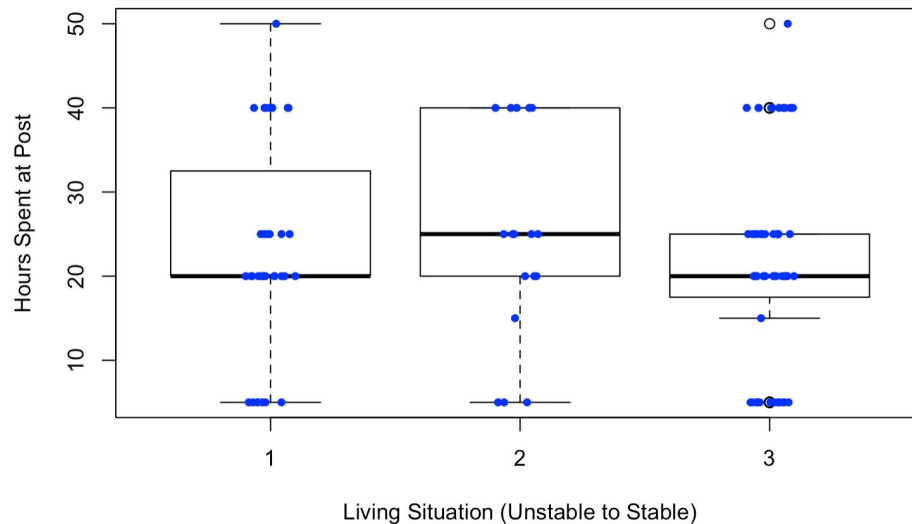
**Summary:** We did **NOT** find a relationship between the given predictor and response.

**Procedure:** In processing the data, we firstly converted both the living arrangement and the hours at post variables to numerical variables. We removed any rows with no response, or unclear responses to either question. Additionally, we changed the value of the "hours spent at post" variable to be the midpoint of the range of hours, i.e. that 20-30 hours would become 25 hours. For the categories of "less than 10" and "more than 50" hours, we changed these values to 10 and 50, respectively. Lastly, we rated living situations on a scale of 1-3 based on the stability of the living situation. See the table below for specific details. If a vendor selected more than one option, we chose to include only the least stable option. This way we could accurately compare the scale of the hours at post, and could actually try to use it to predict the living situation of a worker.

1	2	3
Sleeping outside Sleeping in a car Sleeping in a shelter	Sleeping with friends/family Transitional Housing	Subsidized housing Affordable housing Market rate housing Assisted living facility

To fit the data, we performed linear regression. We started with simple linear regression, and also found fits of higher order polynomials. To analyze the fit, we looked at the p-values of the coefficients (as these are the parameters that indicate whether or not there is a response between the predictor and response). If the p-value is smaller than about 0.05, the predictor is considered statistically significant.

**Results and Discussion:** After trying multiple polynomial fits for the data, we found that there is no relationship between a vendor's living situation and the hours spent at their post.



As we can see, there is no clear visual relationship between the stability of a vendor's living situation with the number of hours spent at their post.

The results of the fits created by linear regression confirm this result: this predictor is not statistically significant in predicting the response.

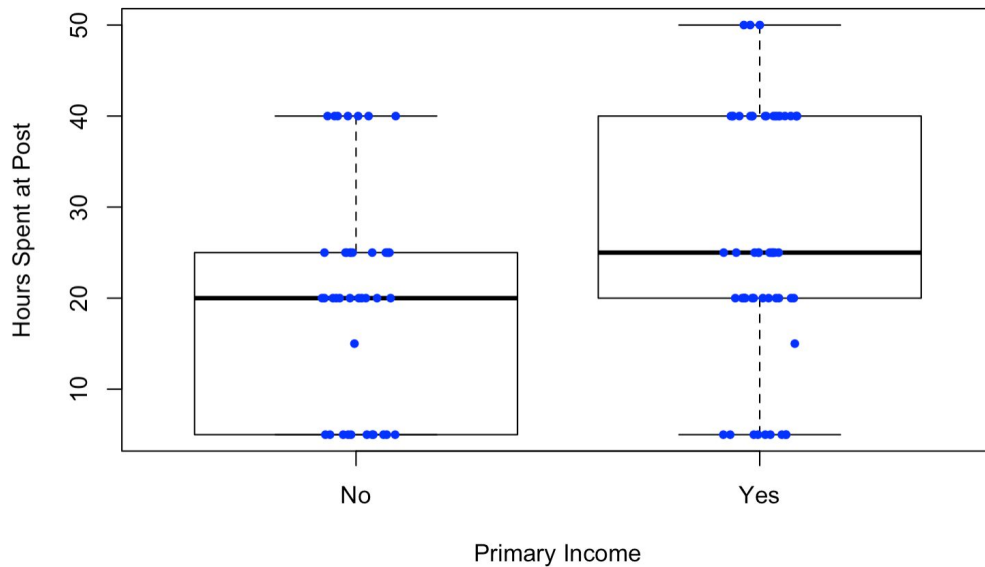
**Further Analysis:** We investigated other predictors that may have effects on the numbers of hours spent at a vendor's post, as well as interaction terms with the living situation predictor.

Predictors we considered but disregarded after analysis:

- Homelessness status
- Age
- Race
- Gender
- Disability
- Earning limit
- Paper familiarity (though vendors who claimed to “Let the paper sell itself” worked fewer hours in general)

Predictors found to be useful:

- Primary income



Visually, there appears to be a relationship between whether or not Real Change is a vendor’s primary source of income and the hours spent at their post. Performing linear regression, we found that there is indeed a significant relationship between these two variables. These results are summarized in the table below.

	<b>NOT Primary source of income</b>	<b>Primary source of income</b>
<b>Average hours spent at post</b>	20.6	26.9

Compared to vendors for whom Real Change is not the primary source of income (who work about 20.6), vendors for whom Real Change is the primary source of income work on average 6.3 hours more.

