MEMORANDUM

To: Alison K. Ventura

From: Team E (Andrew Harlow, Jessica Eng, Dylan Allred, Anastasia Arsky)

Date: May 18, 2022

Re: Predictors of Breastfeeding Among WIC Participants Statistical Analysis

---

The purpose of this memo is to describe the statistical methods and findings from an analysis of your breastfeeding data. I hope that this information helps you address your research project questions:

> *1. "Does participation in a responsive feeding intervention predict greater likelihood of any breastfeeding at 11 months?"*
>
> *2. "What combination of factors assessed during early infancy (between birth and 6 months postpartum) predict likelihood of any breastfeeding at 11 months for low-income mothers of young infants participating in the LA County WIC program?"*

This memo is organized into four sections.

1. The first section, "**Abstract of Key Findings**," presents an overview of key results from our analysis. (page 2)
2. The second section, "**Background and Data**," includes a description of our understanding of your data and of your main statistical question, along with a description of some statistical issues that are important in selecting and then explaining a statistical analysis of your data. (page 3-4)
3. The third section, "**Statistical Methods**," describes an analysis approach for your consideration. (page 5-8)
4. The fourth section, "**Results and Discussion**," describes the results from an analysis for your consideration. (pages 9-11)
5. The fifth section, **"Summary of Key Findings,"** summarizes and discusses the key findings based on the statistical analyses. (page 12)
6. The last section, "**Technical Output,**" includes varied computer output for reference. (pages 13-16)

If you have any additional questions about this work following our consulting meeting today, feel free to contact us at hsmith@calpoly.edu and hglanz@calpoly.edu so that we may set up another meeting to discuss your questions.

I. Abstract of Key Findings:

- Participation in the intervention clinic did not predict greater likelihood of any breastfeeding at 11 months.

- The combination of responsive feeding style and the number of feedings per day predicts an increased likelihood of any breastfeeding at 11 months.

II. Background and Data:

Our understanding is that you want assistance in answering your research questions:

1. "Does participation in a responsive feeding intervention predict greater likelihood of any breastfeeding at 11 months?"

2. "What combination of factors assessed during early infancy (between birth and 6 months postpartum) predict likelihood of any breastfeeding at 11 months for low-income mothers of young infants participating in the LA County WIC program?"

Data Preparation

From what we understand, the variable we are predicting for both research questions is whether the mother is breastfeeding at 11 months or not. To accurately measure this, we subsetted our data to only include the survey responses recorded at 11 months. We also created a new variable called BREASTFEED that essentially converts the given variable CURRENTFED from four categories (breastfeeding only, formula only, formula and breastfeeding, no longer receives either) to two categories (yes breastfeeding or no breastfeeding). More information will be outlined below.

Variables of Interest for Research Question 1

- Treatment: categorical variable representing which clinic the mother participated in (1 → intervention clinic, 0 → control clinic)
- CURRENTFED: categorical variable of type of milk received by infant with 4 levels (1 → breastfeeding only, 2 → formula only, 3 → formula and breastfeeding, 4 → no longer receives either)
- Breastfeed: categorical variable generated from CURRENTFED that represents whether the infant is currently being breastfed or not
  (1 → the infant is being breastfed (breastfeeding only + formula and breastfeeding),
  2 → the infant is not being breastfed (formula only + no longer receives either))

Variables of Interest for Research Question 2

- MOMETHNIC: Mother's ethnicity
  (0 → Non-Hispanic, 1 → Hispanic)
- Primiparous: categorical variable representing number of children a mother has
  (0 → 1 child, 1 → multiple children)
- M6sur: quantitative variable representing infant temperament in the Surgency/Extroversion dimension

- M6neg: quantitative variable representing infant temperament in the Negative Affectivity dimension
- M6eff: quantitative variable representing infant temperament in the Orienting/Self-Regulation Capacity dimension
- Responsive: quantitative variable representing whether the mother used the type of feeding style that is more desirable to the baby or not (possible score range: 1-5)
- Pressure: quantitative variable representing whether the mother used the type of feeding style that is less desirable to the baby or not (possible score range: 1-5)
- NUMFEEDS: quantitative variable representing the number of feedings per day
- WIC_SATISFIED: categorical variable treated as a quantitative variable (because values are on a likert scale) describing satisfaction with WIC experience
- Breastfeed: categorical variable generated from CURRENTFED that represents whether the infant is currently being breastfed or not
  (1 → the infant is being breastfed (breastfeeding only + formula and breastfeeding),
  2 → the infant is not being breastfed (formula only + no longer receives either))

Scope of Inference

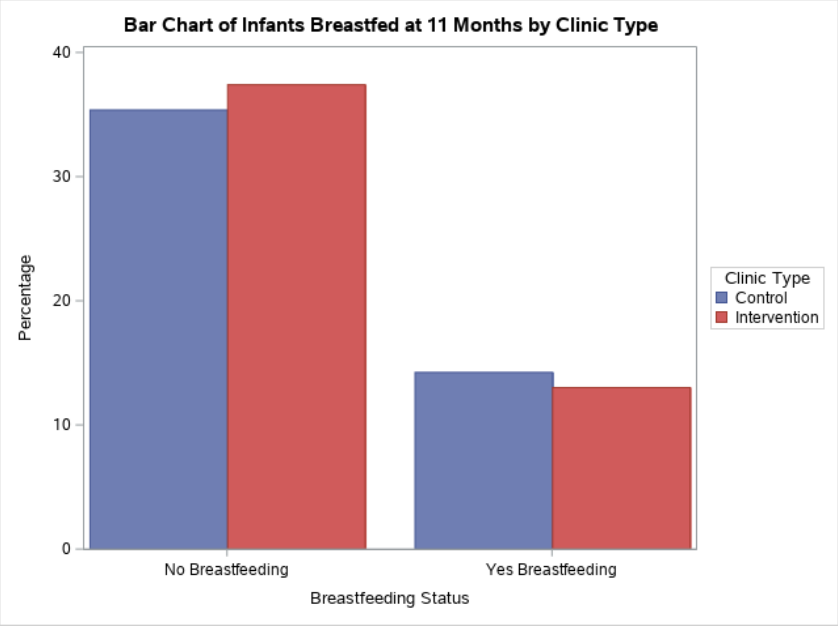
We filtered the data to mothers who were surveyed when their infant was 11 months, so we can only generalize our findings to mothers who are and are not breastfeeding at 11 months. Given that the data were collected from WIC clinics in Southern California our scope of analysis is limited to the region in which the data was collected or regions that have similar population and demographics characteristics.

III. Statistical Methods:

Research Question 1

We understand that your first research question aimed to answer whether participation in the intervention clinic led to a greater likelihood of breastfeeding at 11 months. Before delving deeper into this question, we first wanted to determine if there even was a significant difference between the proportions of breastfeeding mothers participating in the intervention versus control clinics (based on the Breastfeed variable outlined in the previous section). Figure 1 features a bar chart of the percentage of breastfed versus not breastfed infants across the two clinic types. Looking at the bar chart, we can see that the difference in breastfeeding proportions is very slight across the two clinics, though it does seem like participation in the intervention clinic leads to a slightly lower likelihood of breastfeeding at 11 months (the intervention clinic has a lower percentage of breastfeeding mothers and a higher percentage of not breastfeeding mothers compared to the control clinic).

Figure 1. Bar chart of breastfed infants at 11 months by clinic type (control or intervention)



In order to actually analyze whether this difference was significant, we decided to run a Two Proportion z-test. A Two Proportion z-test checks to see if the proportion of two separate populations are equal or not. For this study, the two populations of interest are the participants of the intervention versus control clinics and the proportion of interest is the proportion of mothers still breastfeeding at 11 months in each population.

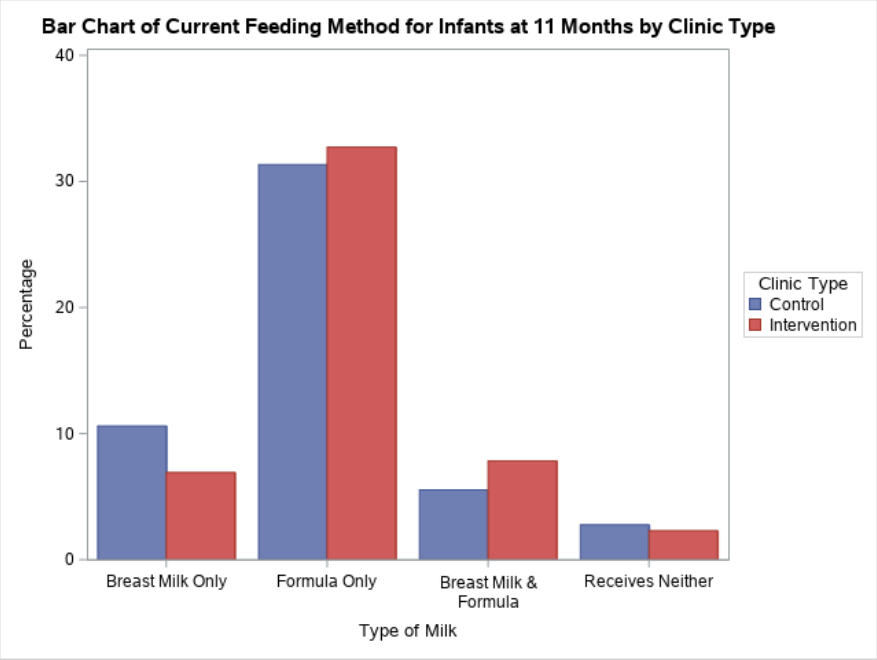
The Two Proportion z-test tests the hypotheses:

Null Hypothesis: The participation in intervention clinics does not have an effect on mothers still breastfeeding at 11 months.

Alternative Hypothesis: The participation in intervention clinics predicts a greater likelihood of mothers still breastfeeding at 11 months.

We also completed a further analysis that tested whether there was a difference across any of the infant feeding methods, not just in breastfeeding specifically, using the variable CURRENTFED. Once again, we first generated a bar chart (Figure 2), and found that, while the differences across CURRENTFED responses were greater than that for Breastfeed, they are not drastically different. It also appears that participating in the intervention clinic encouraged more of the mothers to utilize formula instead of breast milk (Formula Only and Breast Milk & Formula percentages are higher for the intervention clinic while Breast Milk Only percentages are lower compared to the control clinic).

Figure 2. Bar chart of current infant feeding method at 11 months by clinic type (control or intervention)



For this analysis, we used a Chi-Square test of homogeneity, as it determines whether frequency counts are distributed identically across two different populations (intervention clinic vs control clinic).

The Chi-Square test of homogeneity tests the hypotheses:

Null Hypothesis: The distribution of the infant feeding methods at 11 months is the same for both the intervention clinic participants and the control clinic participants.
Alternative Hypothesis: The distribution of the infant feeding methods at 11 months differs across the two populations.

Research Question 2

Furthermore, we understand your second research question is aimed at finding the most important predictors for determining the likelihood of any breastfeeding at 11 months for low-income mothers of young infants participating in the LA County WIC program. Since breastfeeding by mothers at 11 months (the response variable Breastfeed) is binary, meaning there are only two possible values (Yes breastfeeding and No breastfeeding), rather than numerical, we looked to construct a Logistic Regression model. Instead of analyzing the correlation between the response variable and the predictor variables like in Linear Multiple Regression, a Multiple Logistic regression model determines the odds ratios for each predictor quantifying the strength of association between the predictor variable and the likelihood of the event of an infant still being breastfed at 11 months).

In the Multiple Logistic regression model used to investigate your research question, we used the following variables as predictors: MOMETHNIC, Primiparous, m6sur, m6neg, m6eff, Responsive, Pressure, NUMFEEDS, WIC_SATISFIED. During the initial meeting, you discussed that these variables may be of particular interest in predicting the likelihood of a mother still breastfeeding at 11 months, so we have included them in the model. The model also allows for the relationship between responsive feeding style and the probability of mothers still breastfeeding at 11 months to differ depending on NUMFEEDS, which is an interaction between Responsive and NUMFEEDS. We will consider a predictor to be statistically significant if the p-value is less than $\alpha = 0.05$.

We also want to note 63 observations due to missing values in the response and the predictor variables.

Conditions and Assumptions

Before moving forward with the Two-Proportion z-test, we needed to check the model conditions. Each group has at least 10 successes, which can be shown in Table 1 in the Results and Discussion section, so we have a big enough sample size to run this test. However, this test does require simple random sampling which we understand was not used in this data collection process. The implications of this condition not being met is that the results cannot be generalized to the full population of low-income mothers who participated in the LA County WIC program.

Before moving forward with our Logistic Regression model, we needed to check the model conditions. We found no influential outliers to be concerned about and that independence of all observations was satisfied. No issues of multicollinearity (many explanatory variables being correlated with each other) arose in the study. Lastly, our response variable is binary, therefore all of the conditions are met for our model.

IV. Results and Discussion:

Research Question 1

The sample proportions in Table 1 show that 28.69% of mothers in the control clinic group were still breastfeeding at 11 months and 25.81% of mothers in the intervention clinic group were still breastfeeding at 11 months.

Table 1. Numeric summaries of breastfeeding status at 11 months by clinic type (control or intervention)

Sample Proportions and Counts

	Control Clinic	**Intervention Clinic**	**Totals**
Yes Breastfeeding	28.69% (35)	25.81% (32)	27.24% (67)
No Breastfeeding	71.31% (87)	74.19% (92)	72.76% (179)
Totals	49.6% (122)	50.4% (124)	1 (246)

The results from the Two Proportion z-test in Table 2 did not indicate a significant difference in the proportion of mothers breastfeeding between the control clinic group and the intervention clinic group. The test statistic, the difference in proportions between the control and intervention group, was 0.0288, which corresponded to a p-value of 0.6116. This is a very large P-value, so we fail to reject the null hypothesis and thus do not have strong evidence to conclude that there is a difference in the proportion of mothers breastfeeding between the control clinic group and the intervention clinic group.

Table 2. Numeric summary of Two Proportion z-test results

	$\widehat{p_C}$	$\widehat{p_I}$	Test Statistic $= \widehat{p_C} - \widehat{p_I}$	z-score	P-value
Value	0.2869	0.2581	0.0288	0.5078	0.6116

** $\widehat{p_C}$ is the sample proportion of mothers breastfeeding in the control clinic group and $\widehat{p_I}$ is the sample proportion of mothers breastfeeding in the intervention clinic group.

Therefore, participation in the intervention clinic did not predict greater likelihood of any breastfeeding at 11 months.

Furthermore, the sample proportions in Table 3 show that, when looking at infant feeding methods in general instead of just focusing on breastfeeding vs not breastfeeding, the sample distributions are fairly similar across the two populations.

Table 3. Numeric summaries of infant feeding methods at 11 months by clinic type (control or intervention).

Sample Proportions and Counts

	Intervention Clinic	**Control Clinic**	**Totals**
Breast Milk Only	6.91% (15)	10.6% (23)	17.51% (38)
Breast Milk & Formula	7.83% (17)	5.53% (12)	13.36% (29)
Formula Only	32.72% (71)	31.34% (68)	64.06% (139)
Receives Neither	2.3% (5)	2.76% (6)	5.07% (11)
Totals	49.77% (108)	50.2% (109)	1 (217)

**The Chi-Square method removed 12% (29) observations.

Moreover, the results of the Chi-Square test of homogeneity in Table 4 indicate that there are no significant differences between the frequency distributions of infant feeding methods for the intervention versus the control clinics. The corresponding p-value for the Chi-Square (χ^2) statistic was 0.4407, indicating that there is a 44% probability of observing this same data if there really is no difference between the distributions of infant feeding behavior (i.e. the null hypothesis is true). This is a very large P-value, so we fail to reject the null hypothesis and thus do not have strong evidence to conclude that there is a difference in the distribution of infant feeding methods for the two populations (i.e. the participants of the intervention vs the control clinics).

Table 4. Numeric summary of Chi-Square test of homogeneity results

	Statistic	DF	χ^2 statistic	P-value
Value	Chi-Square	3	2.6974	0.4407

Therefore, participating in the intervention clinic does not result in a greater likelihood of breastfeeding, nor does it seem to influence any change in infant feeding method at all.

Research Question 2

Results from fitting the Multiple Logistic regression model in Table 5 indicated that there is a statistically significant association between responsive feeding style and a mother still breastfeeding at 11 months (p-value = 0.0293), after adjusting for MOMETHNIC, Primiparous, m6sur, m6neg, m6eff, Pressure, NUMFEEDS, WIC_SATISFIED, and the interaction between Responsive and NUMFEEDS. For every 1 point increase in the score for responsive feeding style subscale, the expected odds of a mother breastfeeding at 11 months decreases by 0.0790,

after adjusting for the other variables in the model. Therefore, the probability of a mother breastfeeding at 11 months is lower when a mother uses a more responsive feeding style. There is also a statistically significant association between NUMFEEDS and a mother still breastfeeding at 11 months (p-value = 0.0174), after adjusting for the other variables in the model as seen in Table 5. For every additional feeding of breastmilk or formula per day given to an infant, the expected odds of a mother breastfeeding at 11 months decreases by 0.1494, after adjusting for the other variables in the model. Therefore, the likelihood of a mother breastfeeding at 11 months is lower when an infant receives more feedings of breastmilk or formula per day.

Moreover, the model results in Table 5 indicate that the interaction between Responsive and NUMFEEDS is a statistically significant predictor of the likelihood of a mother breastfeeding at 11 months (p-value = 0.0038). When the combination of responsive feeding style and number of feedings per day increases, the expected odds of a mother breastfeeding at 11 months increases by 1.7269, after adjusting for the other variables in the model. Thus, the likelihood of a mother breastfeeding at 11 months is greater when responsive feeding style and number of feedings per day increase together.

Table 5. Numeric summary of Multiple Logistic regression model

Parameters	Estimate	Odds Ratio Estimate	Standard Error	Wald χ^2	P-value
Intercept	6.8167	912.9672	5.6757	1.4425	0.2297
MOMETHNIC 0	0.3806	1.4632	0.2395	2.5246	0.1121
Primiparous 0	0.3713	1.4486	0.2336	2.5254	0.1120
m6sur	-0.1670	0.8462	0.3201	0.2722	0.6018
m6neg	0.3112	1.3651	0.1916	2.6363	0.1044
m6eff	0.0583	1.0600	0.3428	0.0289	0.8651
Responsive	-2.5377	0.0790	1.1643	4.7505	0.0293
Pressure	0.1789	1.1959	0.2965	0.3640	0.5463
NUMFEEDS	-1.9012	0.1494	0.7993	5.6573	0.0174
WIC_SATISFIED	-0.1261	0.8815	0.4632	0.0740	0.7855
Responsive*NUMFEEDS	0.5463	1.7269	0.1885	8.3968	0.0038

**63 observations were deleted due to missing values in the response or predictor variables.

V. Summary of Key Findings:

Research Question 1

Our statistical analyses found that not only is there no significant difference in the proportion of breastfeeding at 11 months between the intervention clinic group and the control clinic group, but there is also no significant difference between the frequency distributions of any of the infant feeding methods (breast milk only, formula only, breast milk and formula, no longer receives either) at 11 months for the intervention vs the control clinics. So, participating in the intervention clinic does not only fail to result in a greater likelihood of breastfeeding, but it does not seem to influence any change in breastfeeding or any other infant feeding method at 11 months whatsoever.

Research Question 2

Our analysis on which combination of factors predict likelihood of any breastfeeding at 11 months found that responsive feeding style, number of feedings per day, and the interaction between these two variables were statistically significant. We found that the probability of a mother breastfeeding at 11 months is lower when a mother uses a more responsive feeding style. The likelihood of a mother breastfeeding at 11 months is also lower when an infant receives more feedings of breastmilk or formula per day. Additionally, the likelihood of a mother breastfeeding at 11 months is greater when responsive feeding style and number of feedings per day increase together.

VI. Technical Output:

Research Question 1

Figure 3. SAS Output for the Two Proportion z-test with frequency, percent, row percent, and column percents

Frequency Percent Row Pct Col Pct

Table of Treatment by breastfeed			
	breastfeed		
Treatment(Treatment)	Yes Breastfeeding	No Breastfeeding	Total
Control	35 14.23 28.69 52.24	87 35.37 71.31 48.60	122 49.59
Intervention	32 13.01 25.81 47.76	92 37.40 74.19 51.40	124 50.41
Total	67 27.24	179 72.76	246 100.00

Figure 4. SAS Output for the Two Proportion z-test risk estimates and p-values

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.2869	0.0409	0.2066	0.3671	0.2086	0.3758
Row 2	0.2581	0.0393	0.1810	0.3351	0.1837	0.3443
Total	0.2724	0.0284	0.2167	0.3280	0.2177	0.3326
Difference	0.0288	0.0568	-0.0824	0.1401		
Difference is (Row 1 - Row 2)						

Risk Difference Test	
H0: P1 - P2 = 0 Wald Method	
Risk Difference	0.0288
ASE (Sample)	0.0568
Z	0.5078
One-sided Pr > Z	0.3058
Two-sided Pr > |Z|	0.6116
Column 1 (breastfeed = Yes Breastfeeding)	

Figure 5. SAS Output for the Chi-Square test of homogeneity frequency and percent counts

Frequency
Percent
Row Pct
Col Pct

Table of Treatment by CURRENTFED					
	CURRENTFED(CURRENTFED)				
Treatment(Treatment)	Breast Milk & Formula	Breast Milk Only	Formula Only	Receives Neither	Total
Control	12 5.53 11.01 41.38	23 10.60 21.10 60.53	68 31.34 62.39 48.92	6 2.76 5.50 54.55	109 50.23
Intervention	17 7.83 15.74 58.62	15 6.91 13.89 39.47	71 32.72 65.74 51.08	5 2.30 4.63 45.45	108 49.77
Total	29 13.36	38 17.51	139 64.06	11 5.07	217 100.00
Frequency Missing = 29					

Figure 6. SAS Output for the Chi-Square test of homogeneity test statistics and p-values

Statistics for Table of Treatment by CURRENTFED

Statistic	DF	Value	Prob
Chi-Square	3	2.6974	0.4407
Likelihood Ratio Chi-Square	3	2.7144	0.4378
Mantel-Haenszel Chi-Square	1	1.1047	0.2932
Phi Coefficient		0.1115	
Contingency Coefficient		0.1108	
Cramer's V		0.1115	

Sample Size = 217
Frequency Missing = 29

WARNING: 12% of the data are missing.

Research Question 2

Figure 7. SAS Output of model specifications for the Logistic Regression Model

The LOGISTIC Procedure

Model Information	
Data Set	BABEZ.WICTEST
Response Variable	breastfeed
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	246
Number of Observations Used	183

Response Profile		
Ordered Value	breastfeed	Total Frequency
1	1	58
2	2	125

Probability modeled is breastfeed=1.

Note: 63 observations were deleted due to missing values for the response or explanatory variables.

Figure 8. SAS Output for categorical variable encoding in the Logistic Regression Model

Class Level Information		
Class	Value	Design Variables
MOMETHNIC	0	1
	1	-1
primiparous	0	1
	1	-1

Figure 9. SAS Output for the global null hypothesis tests for the Logistic Regression Model

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	70.3930	10	<.0001
Score	57.3577	10	<.0001
Wald	39.0596	10	<.0001

Figure 10. SAS Output for the Regression Model maximum likelihood estimates and p-values

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	6.8167	5.6757	1.4425	0.2297
MOMETHNIC	0	1	0.3806	0.2395	2.5246	0.1121
primiparous	0	1	0.3713	0.2336	2.5254	0.1120
m6sur		1	-0.1670	0.3201	0.2722	0.6018
m6neg		1	0.3112	0.1916	2.6363	0.1044
m6eff		1	0.0583	0.3428	0.0289	0.8651
responsive		1	-2.5377	1.1643	4.7505	0.0293
pressure		1	0.1789	0.2965	0.3640	0.5463
NUMFEEDS		1	-1.9012	0.7993	5.6573	0.0174
WIC_SATISFIED		1	-0.1261	0.4632	0.0740	0.7855
responsive*NUMFEEDS		1	0.5463	0.1885	8.3968	0.0038

Figure 10. SAS Output for the Logistic Regression odds ratio estimates and confidence limits

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
MOMETHNIC 0 vs 1	2.141	0.837	5.475
primiparous 0 vs 1	2.101	0.841	5.251
m6sur	0.846	0.452	1.585
m6neg	1.365	0.938	1.987
m6eff	1.060	0.541	2.075
pressure	1.196	0.669	2.138
WIC_SATISFIED	0.882	0.356	2.186