Machine Learning with UC Irvine's Gottman Divorce Predictors Dataset

By Leila Moinpour 2024

Introduction

Through 40+ years of psychotherapy research, Dr. John Gottman developed multiple models to predict marital stability and divorce in couples [2]. In Gottman's book, The Seven Principles for Making Marriage Work, he outlines four negative behaviors that are most likely to predict divorce. These behaviors are known as the "four horsemen" which are contempt for a partner, criticism of a partner's personality, stonewalling, and defensiveness [2]. Through his research, he developed a Divorce Predictors Scale (DPS) survey that therapy practitioners use once trained at the Gottman Institute to predict if a couple will stay married or get a divorce. The survey contains 54 questions, with a 1 - 5 likert scale, intending to measure various aspects of the marital relationship connected to the four horsemen and other related research [3]. This paper intends to evaluate the survey, using machine learning classification models, to see if the survey questions do actually predict if a couple is likely to get divorced.

In running the models, we find that the Gottman survey contains many questions that are strong predictors of divorce. We created two predictive models that produced a 97.5% accuracy rate in prediction based on survey responses, using different subsets of the 54 questions provided.

Data

The DPS data was donated in July 2019 to UC Irvine's ML Repository [5]. Respondents answered 54 relationship questions on a scale of 1 to 5. The dataset contained 170 instances, consisting of 84 (49%) divorced and 86 (51%) married responses. Questions varied in scale types, utilizing frequency questions (e.g. "Never - 1" and "Always - 5") and agreement questions ("Strongly Disagree - 1" and "Strongly Agree - 5"). The target variable was provided, with 2 classes indicating whether or not the respondent later got divorced (e.g., "yes" and "no", or 1 and 0) [5]. Additionally, the likert responses came already normalized from 0 - 4 instead of 1 to 5 and pre-processed with respect to removing NA values. A caveat for this paper is that some of the questions themselves seemed to have some significant grammatical errors. If this study were to be replicated, the data needs to be collected from the Gottman Institute or original researchers directly.

Model 1: Data Pre-processing

In further pre-processing of the data, a correlation based feature selection analysis was done in order to remove features highly correlated with one another. This removed any features with features to target correlations below 0.2 or feature to feature correlations below 0.8. As a result, only 16 survey questions (features) remained in the ML sample we used to perform a logistic regression analysis with.

This was done in order to address multicollinearity between features given that many of the questions might be similar or conceptually related. Given that the questions were created through qualitative means, For example, several questions such as "When discussing with my wife, I usually use expressions such as you alway or you never" and "I can use negative statements about my wife's personality during our discussions" might assess level of criticism present in the relationship in slightly different ways.

Specifically, the DPS questions are as provided in Table 1 for Model 1. The full set of questions is provided in the appendix.

**Table 1: Survey Questions Subsetted from 54 Original Questions**

Q1. When one of our discussions goes in a bad direction, the issue does not extend.
Q6. We don't have time at home as partners.
Q7. We are like two strangers who share the same environment at home rather than family.
Q10. My wife and most of our goals are common.
Q31. I feel aggressive when I argue with my wife.
Q42. When I argue with my wife, it's only snaps and I don't say a word.
Q43. I'm mostly thirsty to calm the environment a little bit.
Q44. Sometimes I think it's good for me to leave home for a while.
Q45. I'd rather stay silent than argue with my wife.
Q46. Even if I'm right in the argument, I'm thirsty not to upset the other side.
Q47. When I argue with my wife, I remain silent because I am afraid of not being able to control my anger.
Q48. I feel right in our discussions.
Q49. I have nothing to do with what I've been accused of.
Q51. I'm not the one who's wrong about problems at home.
Q52. I wouldn't hesitate to tell her about my wife's inadequacy.
Q54. I'm not afraid to tell her about my wife's incompetence.

Figure 1: Sample of the Dataset; n = 170, Features = 16

	Atr1	Atr6	Atr7	Atr10	Atr31	Atr42	Atr43	Atr44	Atr45	Atr46	Atr47	Atr48	Atr49	Atr51	Atr52	Atr54
0	2	0	0	0	1	1	1	2	3	2	1	3	3	2	3	1
1	4	0	0	4	0	2	3	4	2	2	2	3	4	4	4	2
2	2	3	2	2	3	3	2	3	2	3	2	3	1	1	2	2
3	3	3	3	3	2	2	3	2	3	2	2	3	3	3	2	2
4	2	1	0	0	1	2	3	0	2	2	1	2	3	2	2	0

After removing the highly correlated features in the original data set, our new model only retained features to feature correlations under the 0.80 threshold. Features above the threshold are considered redundant. This feature correlation matrix reflects the correlation values of the sample used in the ML classification.

Figure 2: Feature Correlation Matrix of Sample

Feature Correlation Matrix
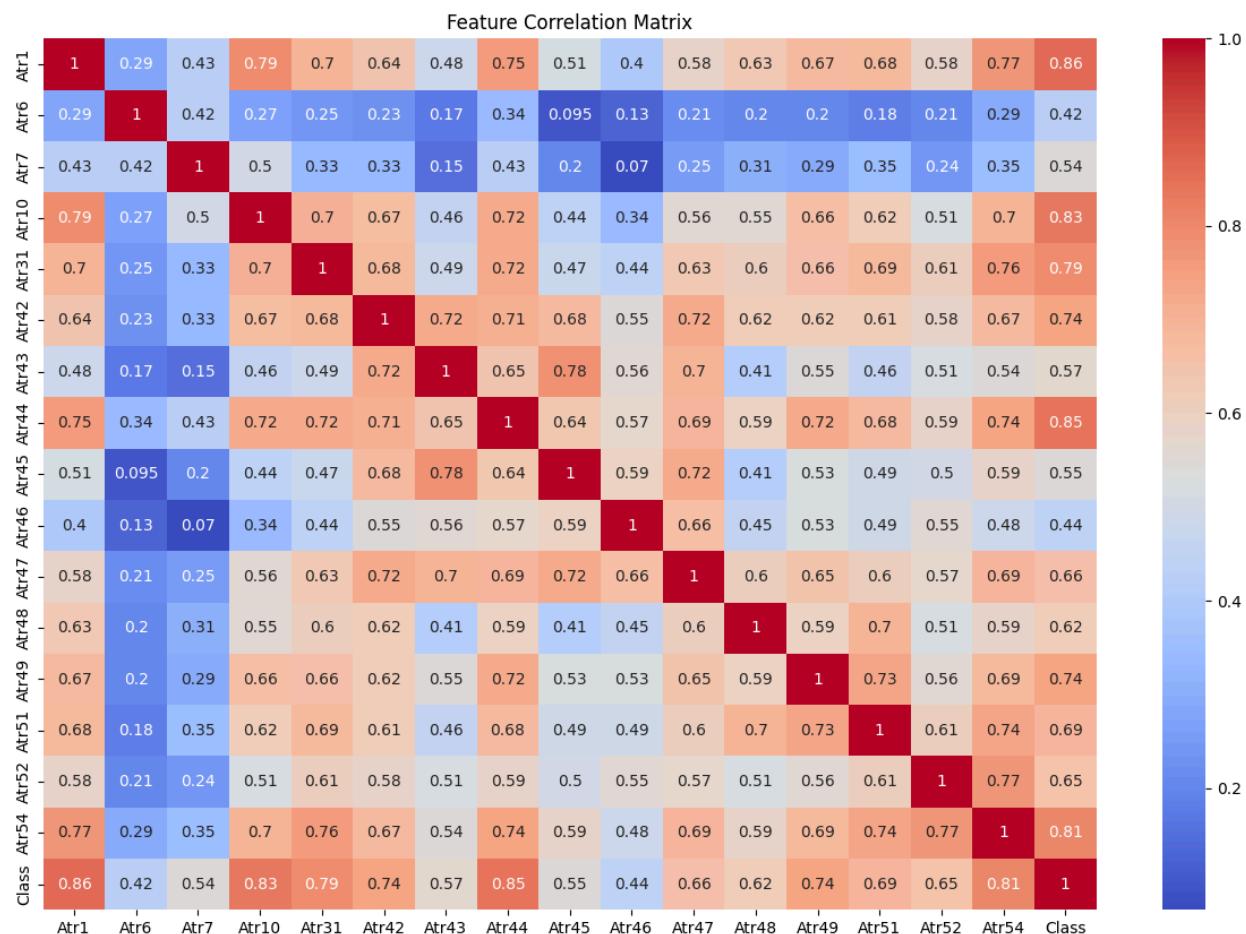
Model 2: Data Pre-processing

Model 2 uses all 54 Features in the dataset, but instead uses Lasso Regularization to deal with multicollinearity between features.

Modeling

Prior to starting the analysis, a Spearman's Rank Correlation was performed to check for a linear relationship between features and the target outcome. It showed a linear relationship, with all 54 features having a statistically significant, high to medium correlation coefficient with respect to the target, ranging from 0.45 to 0.90. Given that a binary classification of the outcome is provided and we have a small sample size of m = 170, utilizing a supervised learning logistic regression model would provide the most direct insight into the relationship between each feature and probability of divorce.

The sample data was normalized to improve the convergence of the algorithm. Min-max normalization was used to convert input features x_i to be within the range 0 to 1. The converted feature x'_i is given by:

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}$$

In code, we used the MinMax Scaler from scikit. After this, we created the training (=0.80) and test sets (=0.20) to prepare for our logistic regression. We used the formula below, and utilized scikit's LogisticRegression package.

$$\hat{y} = h_\theta(x) = \frac{1}{1 + e^{-z}}$$

$$z = \theta_0 + \theta_1 x_1 + \cdots = \boldsymbol{\theta}^T \boldsymbol{x}$$

In running the initial analysis with a simple Logistic Regression, the data was being overfit to the model and had very high coefficient values. As a result, two different approaches were utilized to stabilize the coefficients, creating Models 1 and 2.

Model 1:
Model 1 uses correlation-based feature selection to retain highly predictive features while also removing features highly correlated with other features. This reduces feature redundancy and multicollinearity between features, but allows the model to preserve the most relevant features for analysis. This simplifies the feature set to 16 survey questions prior to the regression analysis. To reduce overfitting, ridge regression (alpha = 0.1) was used to retain all of the features when performing the logistic regression. At a higher alpha level,

Model 2:
In this model, the feature set comprises 54 survey questions. Model 2 relies on lasso regression to both perform regularization and select features that are most relevant to the model by reducing some coefficients to zero. This addresses the high multicollinearity seen in the simple logistic regression prior to experimentation. The alpha used is also 0.01.

Other ML features used to evaluate the model were ROC (Receiver Operating Characteristic) curves, cross validation, confusion matrices, and coefficient analysis.

Results

Our model 1 & 2 results predicted whether or not someone will get a divorce with a 97% accuracy based on the different survey questions with a high AUC. More specifically, the Gottman DPS survey responses do provide enough insight for a prediction to be made.

Table 2: Comparative results of Model 1 & 2

	Model 1	Model 2
Accuracy of model on Test Data	0.970588	0.9705
Training Accuracy	0.9632	0.9852
AUC	1	1
Mean Cross Validation Score	0.9647	0.9529
Standard Deviation of Cross Validation Score	0.07	0.0398
M - number of examples	170	170
N - number of features	16	54
N - number of features after regularization	16	5
Regularization Strength (alpha)	L2 = 0.1	L1 = 0.1

The training accuracy varies between the models but not significantly. The perfect AUC value indicates both models discriminate perfectly between divorce and married classes. However, it suggests that there may be overfitting or data simplicity issues with the models. This requires further investigation with other independent data sets to verify the model. Methodologically, the high AUC should have been addressed through the two different model approaches (e.g. different regularization approaches, and removal of the highly correlated features in model 1). With a smaller standard deviation, model 1 generalizes better across folds.

In comparing both models, model 1 would be preferred given its smaller feature set and smaller standard deviation. It suggests that less survey questions are needed to predict likelihood of divorce. Model 2 retains all 54 questions, which potentially captures other relevant information that was removed in model 1. Both models use different sets of questions as highly predictive indicators in the models.

Figure 3: Model 1 & 2's Confusion Matrix Product Same Results
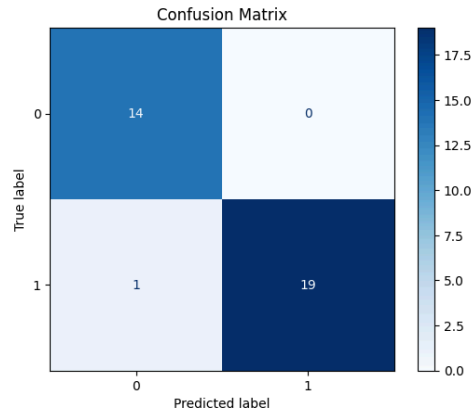
Confusion Matrix

With only one misclassification, the confusion matrix suggests strong performance and supports the AUC value of 1.0.

Figure 4: Model 1 Coefficient Matrix


Feature Importance Based on Coefficients
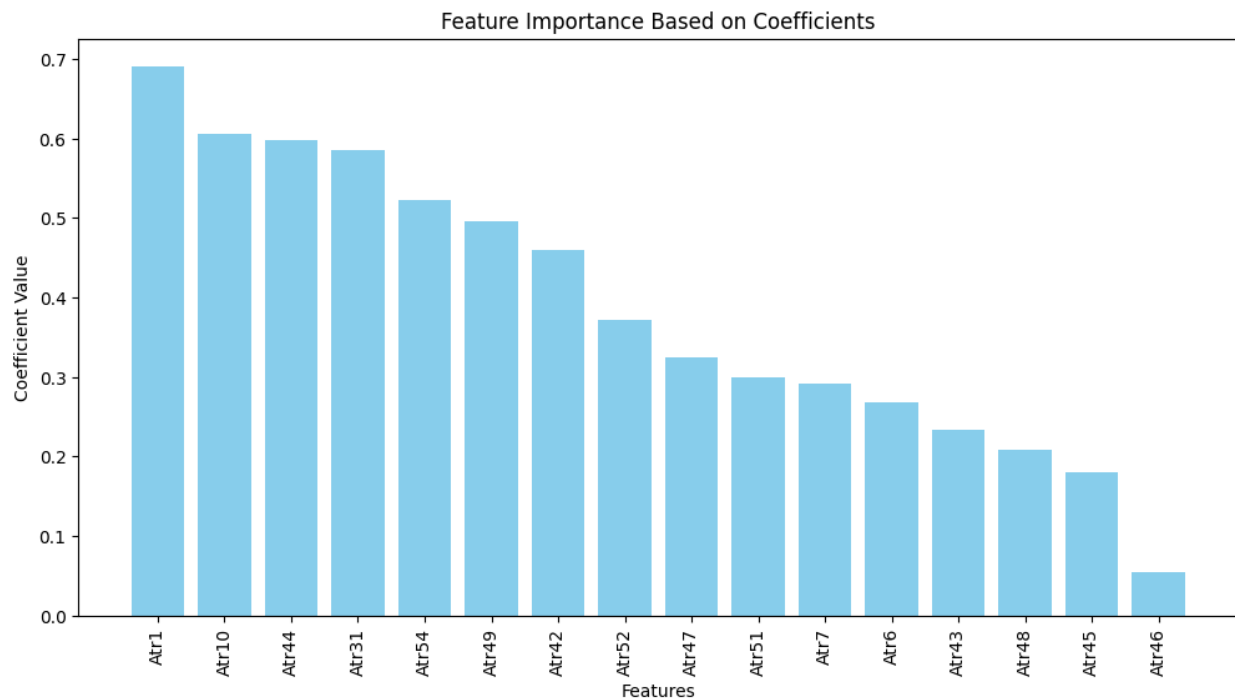
The model identified Atr1 as the most influential predictor, followed by Atr 10, Atr 44, and Atr 31. These features play a significant predictive role in distinguishing between classes in this model. On the other hand, Atr46 has a lower coefficient value, indicating that it is a less influential predictor.

Features ranked in order of highest to lowest coefficients:

1. Atr 1. When one of our discussions goes in a bad direction, the issue does not extend.
2. Atr 10. I enjoy traveling with my wife.

3. Atr 44. When I argue with my wife, it's only snaps and I don't say a word.
4. Atr 31. I know my wife's hopes and wishes.
5. Atr 54. I'm not actually the one who's guilty about what I'm accused of.
6. Atr 49. I have nothing to do with what I've been accused of.
7. Atr 42. When I argue with my wife, it only snaps in and I don't say a word.
8. Atr 52. I wouldn't hesitate to tell her about my wife's inadequacy.
9. Atr 47. When I argue with my wife, I remain silent because I am afraid of not being able to control my anger.
10. Atr 51. I'm not the one who's wrong about problems at home.
11. Atr 7. We are like two strangers who share the same environment at home rather than family.
12. Atr 6. We don't have time at home as partners.
13. Atr 43. I'm mostly thirsty to calm the environment a little bit.
14. Atr 48. I feel right in our discussions.
15. Atr 45. I'd rather stay silent than argue with my wife.
16. Atr 46. Even if I'm right in the argument, I'm thirsty not to upset the other side.

Figure 5: Model 2 Coefficient Matrix
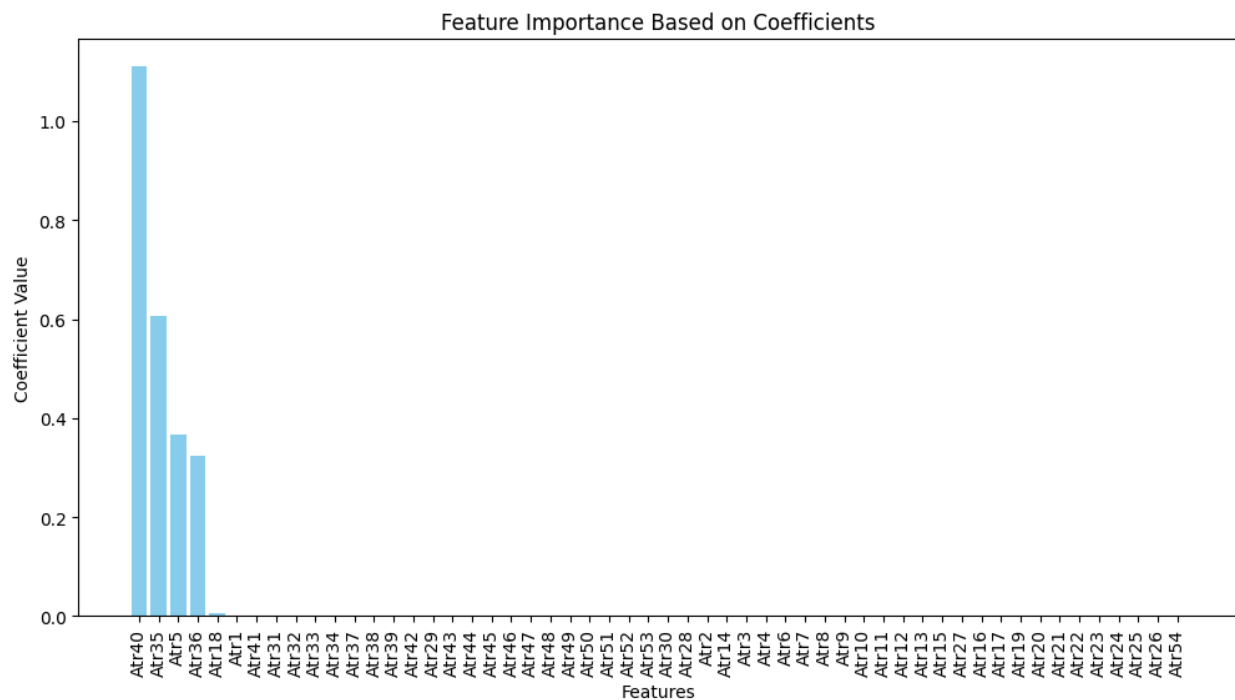
The lasso regularization and logistic regression yielded a model that relies on only 5 features as contributing or influential predictors of class. Atr 40 has a significantly higher coefficient value, suggesting that the feature may simply be strongly correlated with the target variable, have low multicollinearity with other variables, or be a result of model overemphasis on this one feature. Atr 18 seems to contribute little influence to the model, but was significant enough to include.

These features are a set of 5 questions that were influential contributors to the model:
1. Atr 40. We're just starting a fight before I know what's going on.
2. Atr 35. I can insult our discussions.
3. Atr 5. The time I spent with my wife is special for us.
4. Atr 36. I can be humiliating when we argue.

5. Atr 18. My wife and I have similar ideas about how marriage should be.

Discussion

Given the high multicollinearity of features to features within the data set and high predictive results of both models, it suggests that many of the Gottman questions might be significant indicators of divorce. Both models produce entirely separate different groupings of questions with no overlap, and yield a prediction accuracy of 97%.

The implications of this research shows that psychologists can use these survey questions to diagnose if relationships are in need of intervention (i.e. to help couples repair their relationships, etc) and may only need to ask a few questions.

The Gottman Institute has expanded their question subset into 100+ questions for respondents to take in over a 2 hour period [1]. Given the findings, it suggests that the survey length may be reduced significantly, if many questions are intended to help providers predict if a relationship is headed towards divorce.These survey questions can instead be refocused into diagnostics oriented, etc and on giving providers information that enables them to provide better care.

Conclusion

Overall, we can conclude that the Gottman Divorce Predictors Scale (DPS) survey does enable providers to predict if the couple is likely to get divorced or stay married. Originally, the survey questions were created to measure different aspects of the relationship, from compatibility and dynamics to conflict resolution strategies [4]. The ML models propose two different approaches for reducing the survey from 54 questions to 16 and 5 respectively, in order to predict which class the respondent belongs to with 97% accuracy.

In future projects, the Gottman Institute can significantly reduce the number of survey questions being posed to respondents or choose to present alternative questions that would help couples therapists provide better care after reading survey results.

References

[1] "Gottman Connect." Accessed: Dec. 05, 2024. [Online]. Available:
https://gottmanconnect.com/gottman-assessment

[2] "John Gottman - Wikipedia." Accessed: Dec. 05, 2024. [Online]. Available:
https://en.wikipedia.org/wiki/John_Gottman

[3] "Methods Used in the Study and Their Success Rates," ResearchGate. Accessed: Dec. 05, 2024. [Online]. Available:
https://www.researchgate.net/figure/Methods-Used-in-the-Study-and-Their-Success-Rates_tbl1_334170931

[4] "Predicting Marital Stability: An Approach for More Characteristics - The IAFOR Research Archive." Accessed: Dec. 05, 2024. [Online]. Available:
https://papers.iafor.org/submission75975/

[5] "Divorce Predictors data set." UCI Machine Learning Repository, 2019. doi:
10.24432/C53W5P.

Appendix:

Table 1: Full Feature Set

17. When one of our apologies apologizes when our discussions go in a bad direction, the issue does not extend.
18. I know we can ignore our differences, even if things get hard sometimes.
19. When we need it, we can take our discussions with my wife from the beginning and correct it.
20. When I argue with my wife, it will eventually work for me to contact him.
21. The time I spent with my wife is special for us.
22. We don't have time at home as partners.
23. We are like two strangers who share the same environment at home rather than family.
24. I enjoy our holidays with my wife.
25. I enjoy traveling with my wife.
26. My wife and most of our goals are common.
27. I think that one day in the future, when I look back, I see that my wife and I are in harmony with each other.
28. My wife and I have similar values in terms of personal freedom.
29. My husband and I have similar entertainment.
30. Most of our goals for people (children, friends, etc.) are the same.
31. Our dreams of living with my wife are similar and harmonious
32. We're compatible with my wife about what love should be
33. We share the same views with my wife about being happy in your life
34. My wife and I have similar ideas about how marriage should be
35. My wife and I have similar ideas about how roles should be in marriage
36. My wife and I have similar values in trust
37. I know exactly what my wife likes.
38. I know how my wife wants to be taken care of when she's sick.
39. I know my wife's favorite food.
40. I can tell you what kind of stress my wife is facing in her life.
41. I have knowledge of my wife's inner world.
42. I know my wife's basic concerns.
43. I know what my wife's current sources of stress are.
44. I know my wife's hopes and wishes.
45. I know my wife very well.
46. I know my wife's friends and their social relationships.
47. I feel aggressive when I argue with my wife.
48. When discussing this with my wife, I usually use expressions such as you always or you never.
49. I can use negative statements about my wife's personality during our discussions.
50. I can use offensive expressions during our discussions.
51. I can insult our discussions.
52. I can be humiliating when we argue.
53. My argument with my wife is not calm.
54. I hate my wife's way of bringing it up.
55. Fights often occur suddenly.
56. We're just starting a fight before I know what's going on.
57. When I talk to my wife about something, my calm suddenly breaks.

58. When I argue with my wife, it only snaps in and I don't say a word.
59. I'm mostly thirsty to calm the environment a little bit.
60. Sometimes I think it's good for me to leave home for a while.
61. I'd rather stay silent than argue with my wife.
62. Even if I'm right in the argument, I'm thirsty not to upset the other side.
63. When I argue with my wife, I remain silent because I am afraid of not being able to control my anger.
64. I feel right in our discussions.
65. I have nothing to do with what I've been accused of.
66. I'm not actually the one who's guilty about what I'm accused of.
67. I'm not the one who's wrong about problems at home.
68. I wouldn't hesitate to tell her about my wife's inadequacy.
69. When I discuss it, I remind her of my wife's inadequate issues.
70. I'm not afraid to tell her about my wife's incompetence.

Details regarding participants from original paper:

In this study, we did not have access to any demographic data. However, the original paper did, and here is additional information about the participants.

"Of the participants, 84 (49%) were divorced and 86 (51%) were married couples. There were 84 males (49%) and 86 females (51%) in the study group. The ages of the participants ranged from 20 to 63 ($\bar{X}$= 36.04, SD = 9.34). Although the study data were collected from seven different regions of Turkey, the data were predominantly from the Black Sea region (n=79). Of the participants, 74 (43.5%) were married for love, and 96 (56.5%) were married in an arranged marriage. While 127 (74.7%) of the participants had children, 43 (25.3%) had no children. In addition, 18 (10.58%) of the participants were primary school graduate, 15 (8.8%) were secondary school graduate, 33 (19.41%) were high school graduate, 88 (51.76%) were college graduate, and 15 (8.8%) had master's degree. The monthly incomes of the participants were as follows: 34 (20%) individuals had under 2000 TL, 54 (31.76%) had between 2001-3000 TL, 28 (16.47%) had between 3001-4000 TL and 54 (31.76%) individuals had a monthly income over 4000 TL. [3]"