

Statistical Analysis on Factors that Lead to Extramarital Affairs During Marriage

Abstract

This study investigates the correlation between extramarital affairs and various factors including age, years married, and level of religiousness. Utilizing data from a Psychology Today survey conducted in the late 1960s, this research aims to shed light on the dynamics of infidelity within marriages. This research employs regression analysis to explore the relationship between the frequency of extramarital affairs and explanatory variables such as gender, age, years of marriage, level of religiousness, and level of education. Findings suggest a significant relationship between age, years married, and religiousness with the number of extramarital affairs. However, the study faces challenges with meeting the assumptions of linear regression due to issues of non-normality and correlated residuals. As a result, a Proportional Odds (PO) model is employed as a more suitable alternative, which validates the adequacy of the model. Although the study provides valuable insights into the factors influencing extramarital affairs, limitations include the outdated dataset and the need for further transformation to meet regression assumptions. Future research could benefit from using more recent data and exploring evolving societal norms surrounding marriage and infidelity. Additionally, comparing datasets from different time periods could illuminate changes in the impact of factors, such as gender, on extramarital affairs.

Introduction

While monogamy stands as the most common marital practice around the world, the prevalence of infidelity has remained throughout time and is a contributing factor to half of all first marriages in the US ending in divorce. According to the June issue of *Marriage and Divorce*, 70% of the U.S. population in 1977 reported having engaged in an affair while married (Greeley 1994). This alarming statistic remains high to this day. A study conducted by the National Library of Medicine in 2013 revealed that 60% of divorced couples in the US cited infidelity as a reason for their divorce (Rokach A, Chan SH 2023).

The unfortunate reality that infidelity in marriage contributes to divorce in the United States has led to a lot of skepticism and uncertainty for Americans wondering if marriage is the right choice for them. Since one of the most frequently cited causes of a broken marriage is infidelity, it is useful for people to know the consequences of having a partner engage in extramarital affairs. Understanding the factors that are correlated with extramarital affairs will be helpful to those involved in the marriage and divorce process. This may include therapists, family lawyers, and the married individual themselves.

Past literature studies have devoted efforts to exploring the factors leading up to cases of infidelity, wherein factors such as religiousness, educational level, and income are considered (Knopp 2017). However, few studies have explored the impact that the number of children and years in marriage might have and how these variables may correlate with cases of extramarital affairs. Thus, this study seeks to fill the gap in comprehensive research on reasons for engaging in extramarital affairs. By conducting regression analysis on the number of extramarital affairs, this study hopes to analyze the connection between years of marriage, gender, age, level of religiousness, education level, number of children, and individuals' engagement in extramarital affairs.

Methods

This study employs cross-sectional infidelity data, collected from a survey conducted by Psychology Today. Psychology Today is a behavior science magazine founded in 1967 to bring psychology to the masses. In 1969, a questionnaire was published in Psychology Today and readers were requested to mail in their answers. The data were part of a general survey about sex to gain insight into the sexual tendencies of the magazine's readers as well as US society at the time. Around 20,000 responses were received, and approximately 2,000 responses were transferred onto magnetic tape.

The variables defined in this set are the number of occurrences of extramarital sexual intercourse, the response variable, and explanatory variables that include gender, age, years of marriage, parental status, level of religiousness, level of education, level of occupation (based on Hollingshead classification), and self-rating of marriage. For the level of religiousness, the scale goes from 1-5, where 1 indicates no religiousness and 5 indicates strong religiousness. For education level, the scale goes from 9-20: 9 indicates grade school, 12 indicates high school graduate, 14 indicates college, 16 indicates college graduate, 17 indicates graduate work, 18 indicates master's degree, and 20 indicates Ph.D., M.D., or another advanced degree. For occupation level, the scale is based on Hollingshead classification in reverse: 1 indicates those dependent on welfare with no regular occupation, 2 indicates unskilled workers, 3 indicates machine operators and semi-skilled workers, 4 indicates craftsmen and skilled laborers, 5 indicates clerical and sales workers, 6 indicates technicians and semi-professionals, and 7 indicates small business owners (businesses valued at \$50,000-\$70,000) and managers (Nathan Kline Institute). For self-rating of marriage, the scale is based on self-reported data: 1 implies a very unhappy state, while 5 indicates a very happy one. Specifically, the ratings are defined as follows: 1 = very unhappy, 2 = somewhat unhappy, 3 = average, 4 = happier than average, and 5 = very happy.

For this dataset, we will use the number of extramarital affairs as our response variable and gender, age, years of marriage, level of religiousness, and level of education as our explanatory variables

Analysis

The response variable, the number of instances of extramarital sexual intercourse, is numerical and nonbinary.

It is believed that years married and religiousness jointly have a different impact on the number of affairs than each one has individually. To test the interaction between these variables, an interaction plot was generated and hypothesis tests were conducted. The interaction plot revealed an interaction between years married and religiousness as the lines intersected. Since the p-value of the interaction term was 0 and lower than any reasonable alpha, there was sufficient evidence to believe interaction was present.

Prior to checking conditions for inference, variable selection was conducted using backward selection. This study started with a full model including all five predictor variables and then iteratively removed the least significant variables using the AIC (Akaike information criterion) method. The results concluded that removing gender and education resulted in the smallest AIC. Hence a reduced model containing the variables religiousness, years married, and age was selected, with the interaction terms being dropped due to the AIC procedure. In addition, three reduced models were created to ensure robustness and to check if these explanatory variables were significant. In all cases, the p-value was essentially 0, providing evidence that each predictor variable has a relationship with the response variable. This indicates that age, years married, and religiousness should stay in the fitted reduced model. This aligns with the backward selection principle employed in this study, which dictates that no further variables should be removed from the model after the initial elimination. Accordingly, this study reaffirms the use of the complete models for the subsequent analyses.

Figure 1 illustrates the relationship between age and the number of affairs and demonstrates an upward trend in the regression line. Despite the slight increase in affairs as age increases, the scattered data points suggest that the correlation could be weak. The religiousness-affairs plot (Figure 2) presents a downward trend, implying that higher religiousness correlates with fewer affairs. Nevertheless, the scattered data points suggest that the variability in the number of affairs cannot be explained solely by one's level of religiousness. Additionally, with a more pronounced upward trend, Figure 3 indicates a stronger positive relationship between the length of an individual's marriage and the number of affairs. Lastly, the interaction plot between years of marriage and religiousness levels (Figure 4) shows varying slopes for different levels of religiousness, suggesting that the effect of years married on the number of affairs is not uniform across religiousness levels.

Plotting the linear model's fitted values versus the residual values, it was revealed that there were significant trends in the scatterplot which suggests that the variance of the error term is not constant. While we attempted to transform the response variable, the scatterplot still showed strong trends in the data. Thus, there are unequal variances among the residuals. For testing normality, the Shapiro-Wilcox hypothesis test demonstrated that there was evidence of the data having non-normality. The histogram was not a bell-curve shape and was highly skewed to the right. The plot also showed a lack of normality for the residual values in the complete model. For testing independence, the Durbin-Watson hypothesis test was conducted, revealing that the residuals are correlated, and thus not independent. This finding suggests a violation of the independence assumption quintessential for the linear regression, affecting the validation of the analysis. Therefore, this multilinear regression model fails to meet the four model assumptions.

In light of these challenges with the multilinear regression model, this study utilized a viable alternative approach using the Proportional Odds (PO) model to further the robustness of the analytical

framework of this research. After implementing the PO model, this study failed to reject the null hypothesis, indicating that the model is adequate. The PO model's adequacy accordingly implies the satisfaction of the underlying assumptions of linear regressions, the sufficiency of the dataset, and the robustness of the model fit.

Based on the above information, the PO model should be used for our analysis. The equation for the final PO model is shown below:

$$P(y \leq j) = \frac{\exp[-0.11490012 + 0.03676774 \cdot a - 0.13536542 \cdot b + 0.38403352 \cdot c]}{1 + \exp[-0.11490012 + 0.03676774 \cdot a - 0.13536542 \cdot b + 0.38403352 \cdot c]}$$

In this model, 'y' represents the likelihood of engaging in an extramarital affair. The predictor variables include 'a', which represents age; 'b', which indicates the number of years of marriage; and 'c', which reflects the individual's level of religiousness.

Discussions and Conclusion

Based on our results and analysis, our research question was answered when we learned that there is a relationship between the number of extramarital affairs and the factors of age, years married, and level of religiousness for individuals. The PO model concludes that on average, the log odds of being in category j or below versus being above category j will increase by 0.03677 for each additional year in age. Furthermore, the log odds of being in category j or below versus being above category j will decrease by 0.13537 on average for every additional year of marriage. Finally, the log odds of being in category j or below versus being above category j will increase by 0.38403 on average for each additional religiousness rating.

Regarding limitations, there is room for improvement in terms of the year that the data was collected. Rather than choosing a dataset from 1969, a survey could be conducted during the 21st century where marital expectations and factors have likely changed. Given the time in which the data was collected, it was interesting that gender ended up being removed during variable selection due to lack of significance. It may be intriguing to compare a more recent dataset with the one in this study to analyze how the impact of gender differs over time.

Moreover, the fact that the data were collected through a magazine survey and required readers to mail in may introduce selection biases. That is, this approach might have skewed the sample and results toward the readers of Psychology Today, especially the ones who were inclined to respond to the survey and had access to resources necessary to mail. Accordingly, the results might not be representative of the general U.S. population. Additionally, with only 10% of the responses being coded onto tape, biases might arise if the subset was not randomly selected. If done manually, the process of transferring data to magnetic tapes might have introduced inconsistencies and errors.

Since the four assumptions for linear regression were not met, we decided to use the PO model, a more appropriate model, for our analysis of the dataset. In conclusion, we found this experiment to be thought-provoking, especially with an interesting dataset about marital status. Since the dataset was taken from the 1960s, there was inquiry about how different the significant factors would be for the number of extramarital affairs compared to the 2000s. All in all, the factors were similar to what we expected would affect the chance of affairs during marriage.

Bibliography

Fair, R. C. (1978). *A theory of extramarital affairs*. *Journal of Political Economy*, 86(1), 45–61. <https://doi.org/10.1086/260646>

Fair's extramarital affairs data. (2023, May 6). Kaggle. <https://www.kaggle.com/datasets/utkarshx27/fairs-extramarital-affairs-data/data>

Greeley, A. M. (1994). Marital infidelity. *Society*, 31(4), 9–13. <https://doi.org/10.1007/bf02693241>

Knopp, K., Scott, S. B., Ritchie, L. L., Rhoades, G. K., Markman, H. J., & Stanley, S. M. (2017). Once a cheater, always a cheater? Serial infidelity across subsequent relationships. *Archives of Sexual Behavior*, 46(8), 2301–2311. <https://doi.org/10.1007/s10508-017-1018-1>

R: Data and Examples from Greene (2003). (n.d.). <https://search.r-project.org/CRAN/refmans/AER/html/Greene2003.html>

Rokach, A., & Chan, S. H. (2023). Love and Infidelity: Causes and consequences. *International Journal of Environmental Research and Public Health/International Journal of Environmental Research and Public Health*, 20(5), 3904. <https://doi.org/10.3390/ijerph20053904>

Appendix

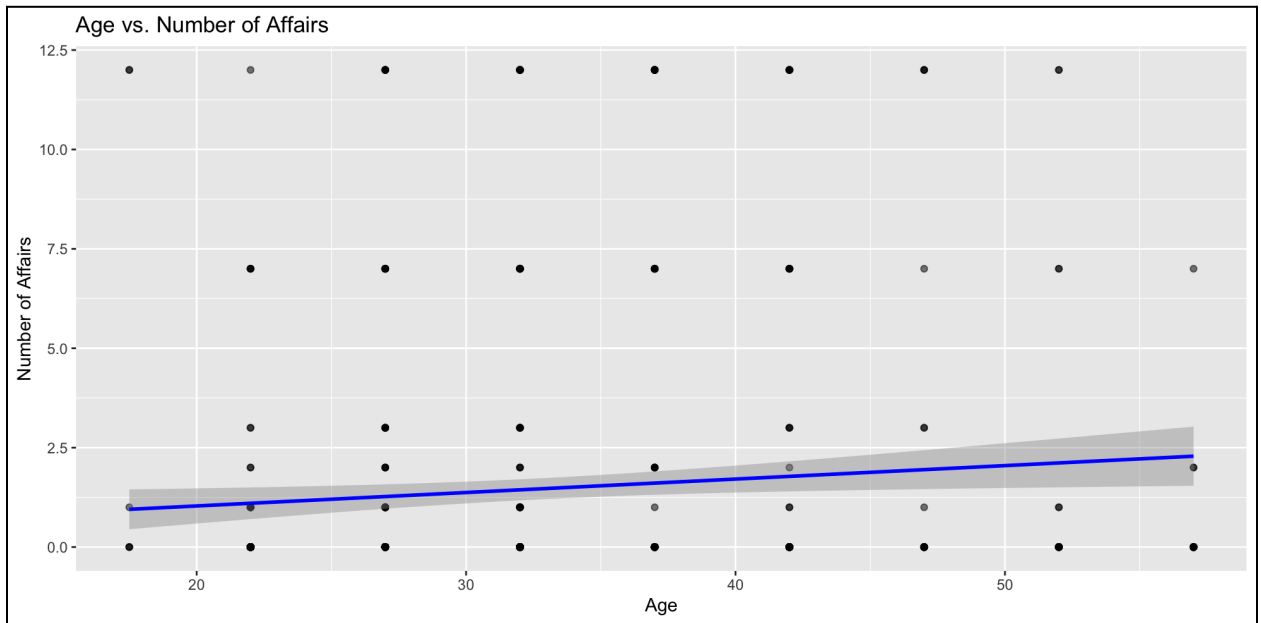


Figure 1

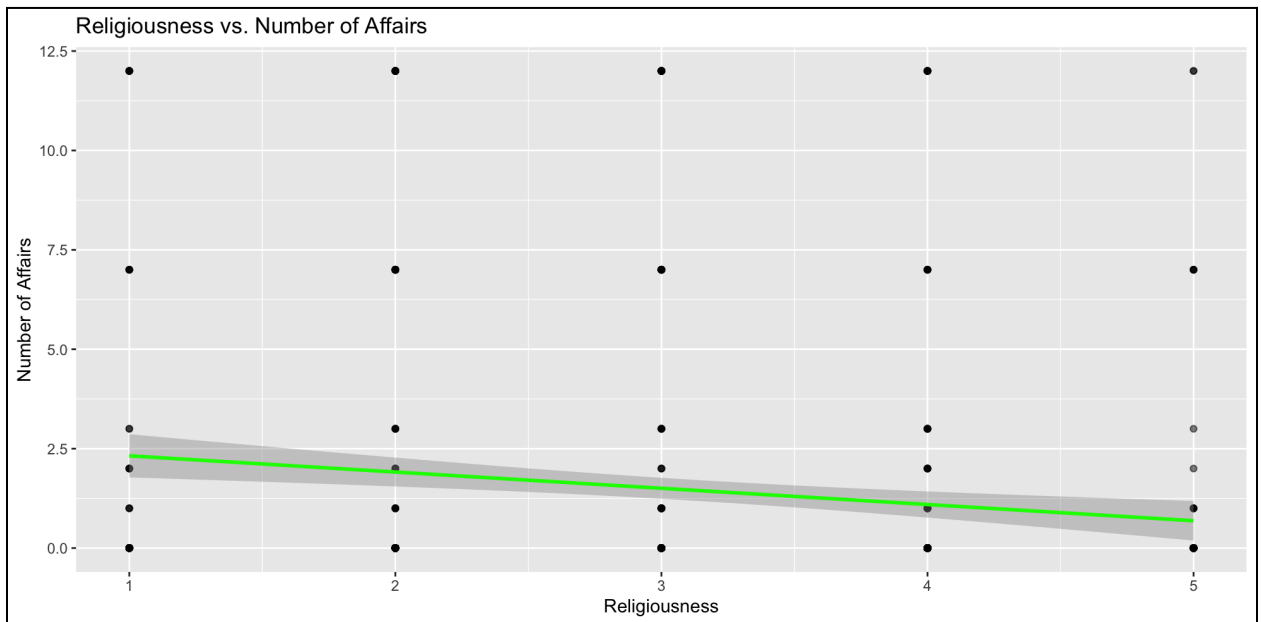


Figure 2

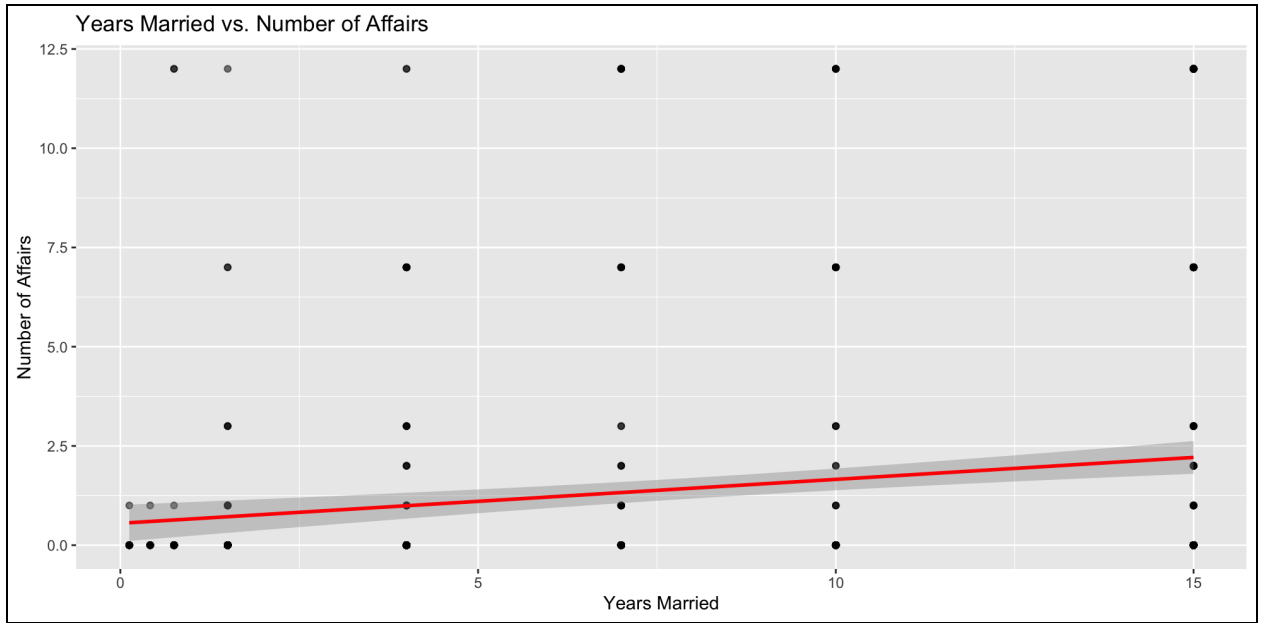


Figure 3

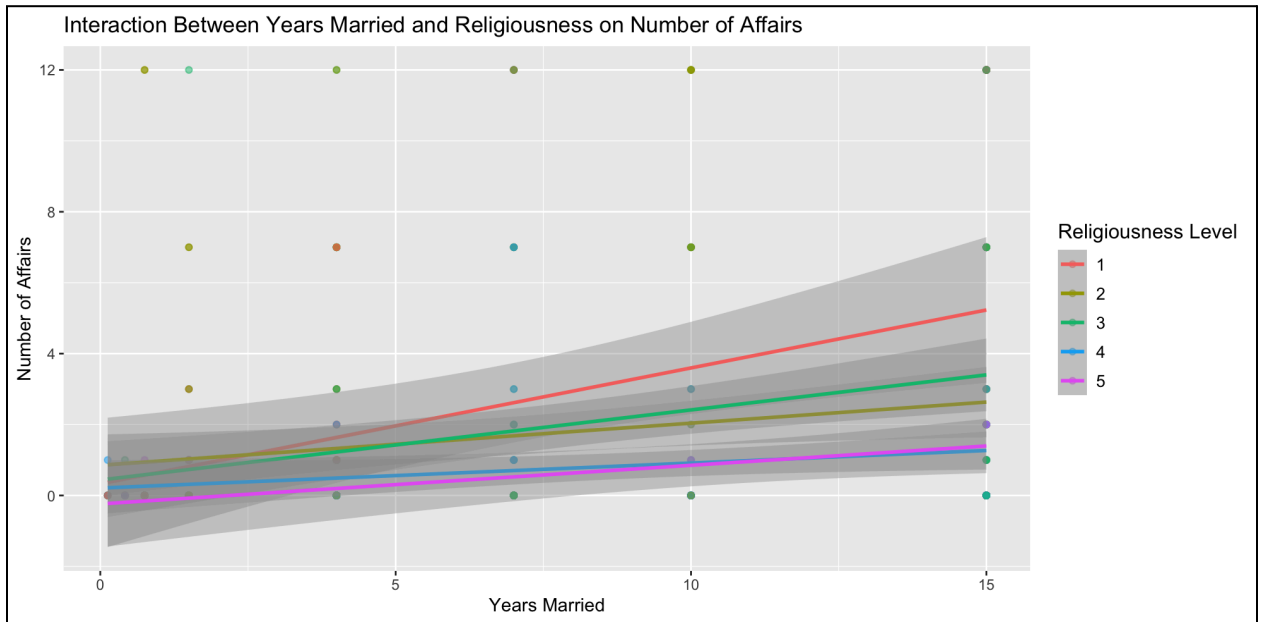


Figure 4