Dacanay, Jordan

Lansangan, Romand

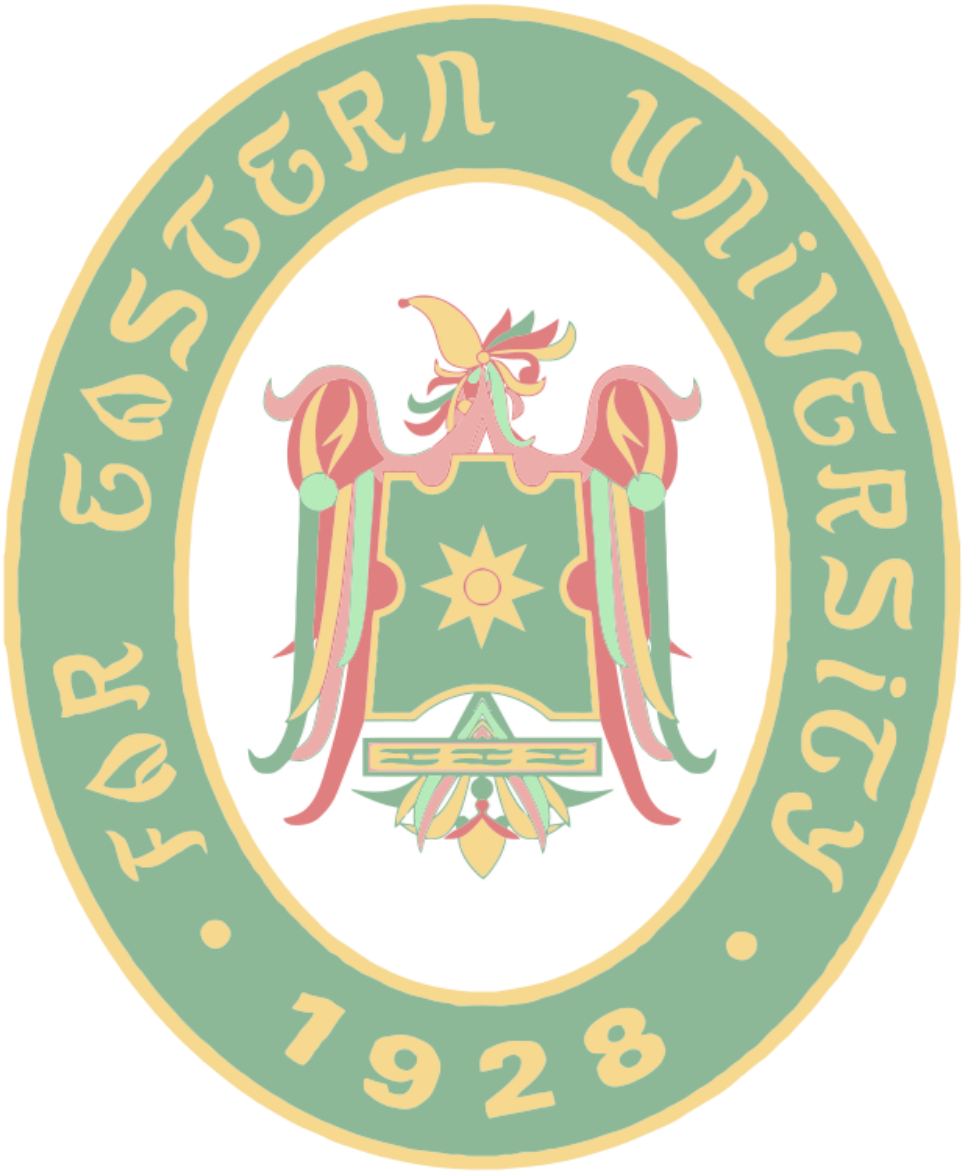
Ramilo, Zion

Analyzing Reproductive Health and Behavior Patterns Using NSFG 2022–2023

DSC 1105 Summative Assessment 2

May 20, 2025

3rd Year Applied Mathematics (Data Science)



Introduction

This report analyzes reproductive health and behavior patterns among female respondents using data from the National Survey of Family Growth (NSFG) 2022-2023. The goal of the said report is analyzing relationship and nature of (selected) variables inside the named dataset. Using different statistical techniques like Posson regression, contingency tables and categorical response modeling, the modelers was able to understand and articulate the complexity of the relationship between different behavioral patterns among female respondents.

Dataset Description

The data for this analysis is drawn from the female respondent subset of the National Survey of Family Growth (NSFG). The NSFG is a national United States based representational survey from National Center for Health Statistics (2024) done with 5586 female respondents to collect 1912 behaviors with regards to their lives, marriages, divorces, pregnancies, contraception, and general health. The dataset along with the male counterpart can be found on the website of U.S Centers for Disease Control and Prevention.

The specific variables used will differ for each part of the report.

Statistical Analyses

Variables to be used:

Original Variable	Description	Type	New Name
OPPLIFENUM	Number of opposite-sex partners in lifetime for all types of sex (top-coded).	Count Variable	sex_partners
GRFSTSX	Grade Respondent Was in at First Sexual Intercourse (bottom-coded).	Ordered categorical	first_sex
MANRASDU	Man Respondent thinks of as raised her during teens	Nominal categorical	man_guardian
WOMRASDU	Woman Respondent thinks of as raised her during teens.	Nominal categorical	woman_guardian
LMARSTAT	Respondent's legal marital status relative to opposite-sex spouses.	Demographics	marital_status

Table 1.a: Variables to be used for statistical analyses

```
(5586, 5)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5586 entries, 0 to 5585
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   sex_partners    5586 non-null   int64
 1   first_sex       5586 non-null   object
 2   man_guardian    5586 non-null   object
 3   woman_guardian  5586 non-null   object
 4   marital_status  5586 non-null   object
dtypes: int64(1), object(4)
memory usage: 218.3+ KB
```

Table 1.b: General Information on the Cleaned Data

Table 1.b shows the general information of the dataset after cleaning where it can be observed that we have the variables of the number of sex partners, the stage where they had any sexual experience, what mother or father figure they have in their household, and lastly the marital status of the respondent.

Table 1.b also shows that there are a total of five thousand five hundred eighty-six observations within the data where it can be seen that there are no null entries which implies that our dataset is complete and does not suffer from missing data issues, thereby ensuring the reliability of any statistical analysis.

Within this data, it is determined that the number of sex partners shall be the dependent variable being evaluated within our statistical analysis, while other variables that are identified to have a data type of object shall become our independent variables.

```
sex_partners: [  1   0   9   8   7   4  15  13   5 999 998   3   6   2  10  11  27  30
  50  25  18  22  12  20  23  14  44  26  17  33  21  40  35  16  19  37
  32  36  24  31  41  45  28  47  34  43  29  38]

first_sex: ['Inapplicable' '10-grade' '11-grade' '9-grade' 'Not-school' '7-grade'
 '3-college' '8-grade' '12-grade' '6-grade-less' 'Refused' '1-college'
 "Don't know" '2-college' '4-college']

man_guardian: ['With both parents' 'No father figure' 'Stepfather' 'Bio father'
 'Refused' 'Other father figure']

woman_guardian: ['With both parents' 'Bio mother' 'Other mother figure' 'Refused'
 'No mother figure' "Don't know"]

marital_status: ['Never married' 'Married' 'Divorced/annulled' 'Separated' 'Refused'
 'Widowed' "Don't Know"]
```

Figure 1: Unique Values per Variable

Figure 1 shows the unique values per variable where it can be observed that within our independent variable, it is composed mostly of discrete values. First-sex experience has thirteen

unique values, Male and Female guardian has six unique variables, and lastly, Marital Status has seven unique values.

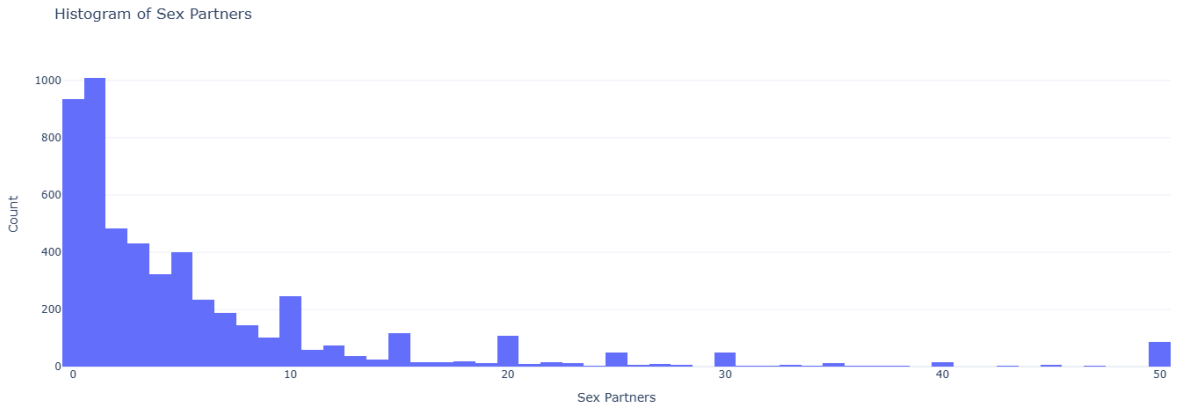


Figure 2: Histogram Plot of Sex Partners

The plot shows that the number of Sex Partners takes on a distribution that is skewed to the right, which implies most of the values present within the data set take a smaller value. In addition, the data has values like 999 or 998 which was removed out of the data set under the grounds that this was suspicious enough of an outlier that would constitute its removal.

Figure 2 shows that the data does not pass the assumption of normality that is needed for powerful statistical analysis. In this case, the transformation of the data set shall be applied. The statisticians shall be using the Yeo-Johnson transformation to address the presence of zero values within the data and ensure that our data is primed for statistical analyses.

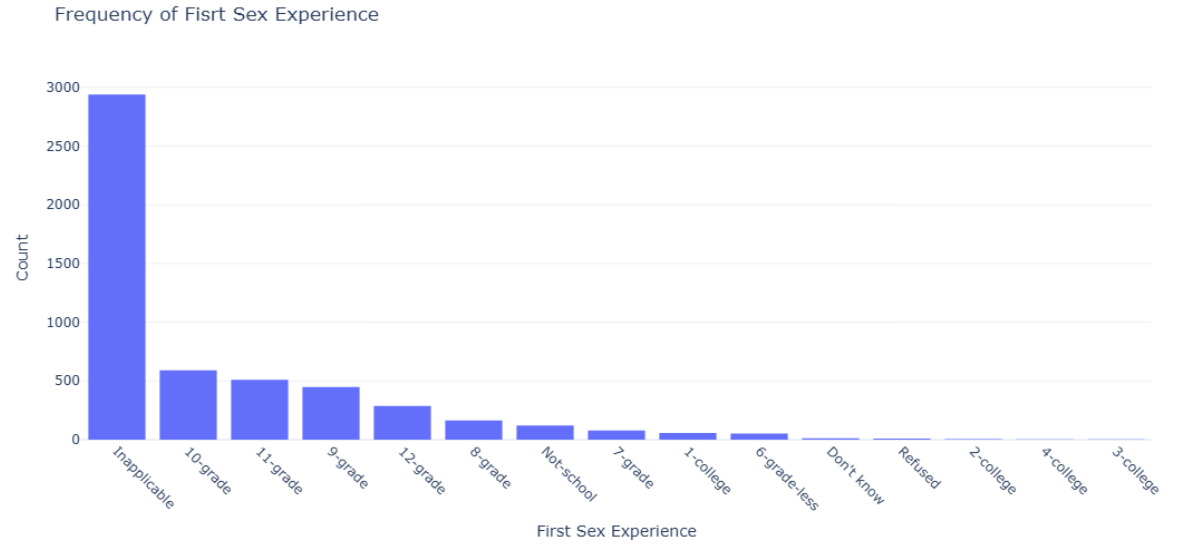


Figure 3: Bar plot of the frequency of the First Sex Experience

Figure 3 shows the distribution of the frequency of the observations for each category of the first sex experience, where within the figure it is shown that more observations were

recorded that it was inapplicable for them, with grade ten being the second most, and eleventh grade being the third most. Furthermore, it has been identified that there are less respondents of the category of sophomore college and above.

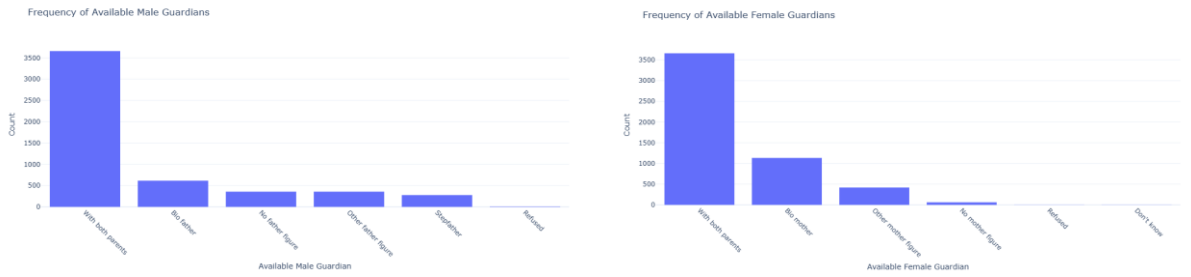


Figure 4: Frequency of Male Guardian Categories (Left) and Frequency of Female Guardian Categories (Right)

Figure 4 shows the distribution of the respondents about the presence of male (left) and female (right) guardians. Where Figure 4 shows that on both male and female guardians most of the respondents have both of their presents present within their lives whilst, for the male guardians the least refused to say, and the least for the female guardian shows that there is a low proportion of respondents either don't know or refused to say.

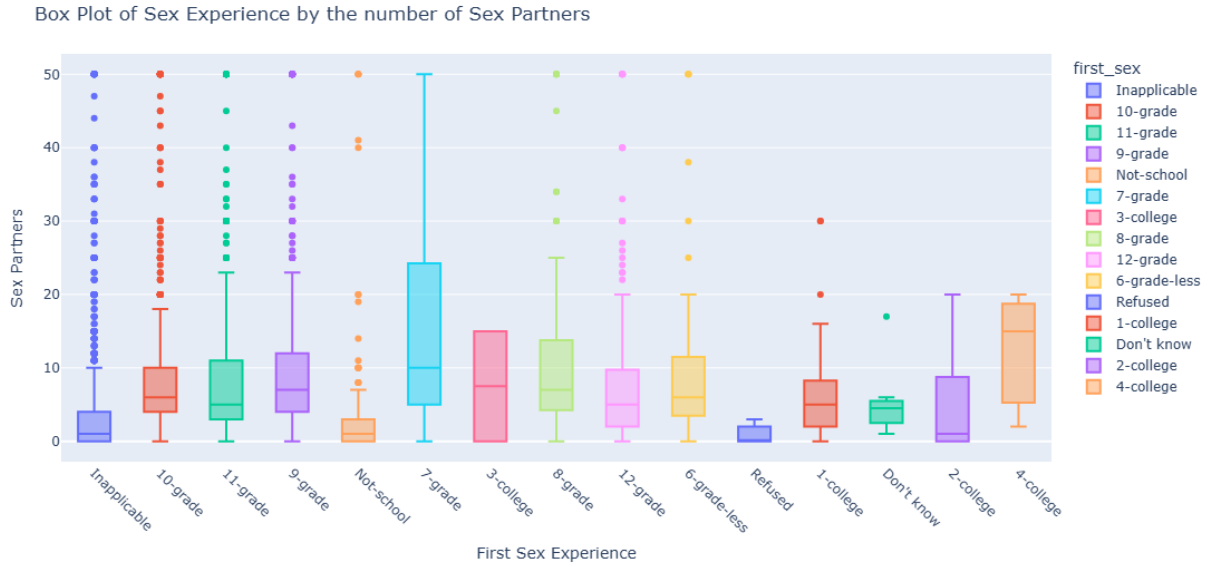


Figure 5: Box plot of First Sex Experience by the number of Sex Partners

Figure 5 shows the distribution of the number of sex partners for every single category of first sex experience, whereas derived from Figure 1 most of the distribution lies within lower values, which can be seen by their interquartile ranges of each category. Furthermore, the distribution for each different category shows that there is much variation where some data is more spread than others.

Figure 5 also shows that there are a lot of outliers for each first-sex experience category which can be a big source that can cause type II or type I errors such that there would be a large probability that it will produce a false positive or false negative. Moreover, another that would

be a cause for type I and II errors is the few samples that are considered for each group, this is an issue as ANOVA requires a certain threshold for each between-group value to enable the model to perform accurately. To mitigate this the statisticians reduced the data into the following: “College”, “Senior Highschool”, “Highschool”, “Middle School”, “Inapplicable”, “Refused”, and “Don’t know”. In addition to this the observations that were identified to take on a value of either “Refused” or “Don’t know” were removed due to the few observations which would not be ideal for one-way ANOVA testing.

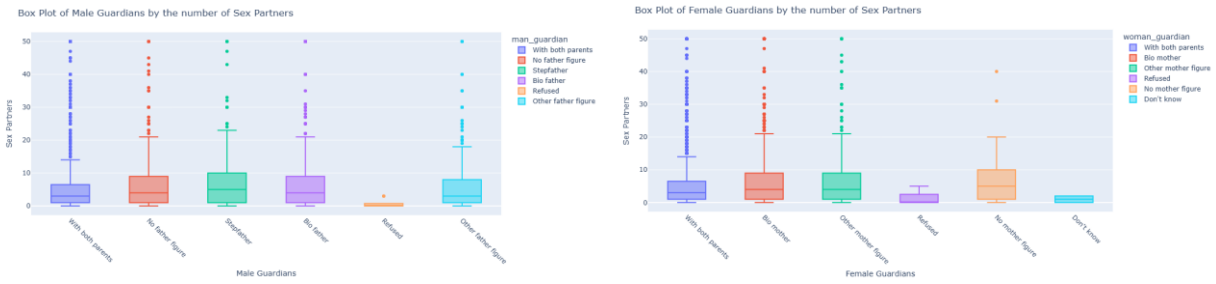


Figure 6: Box plot of Male Guardian Categories (Left) and Box plot of Female Guardian Categories (Right) by the number of Sex Partners

Figure 6 shows the distribution of the respondents within each variable for male and female guardians where it can be seen that there is still a presence of outliers within each value, where four of the categorical values for both the male and female guardians have outliers. In addition, similar to the distribution for the first-sex experience, it shows a distribution that leans more toward lower values.

Within Figure 6 it has been identified that there are categorical values that have fewer observations which would not be ideal for one-way ANOVA. In order to mitigate this problem, these observations were removed from the main dataset.

	sum_sq	df	F	PR(>F)
C(first_sex_combined)	850.062991	5.0	361.963594	0.0
Residual	2466.370013	5251.0	NaN	NaN

Table 2: One-Way ANOVA Test for First-Sex Experience

A one-way ANOVA was conducted to examine the effect of first-sex experience on the dependent variable. There was a statistically significant effect of group, $F(5, 5251) = 361.96$, $p < .001$.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
College	Highschool	0.2971	0.0072	0.0522	0.5419	True
College	Inapplicable	-0.578	0.0	-0.8194	-0.3367	True
College	Middle School & Below	0.294	0.1855	-0.067	0.6551	False
College	Not-school	-0.5228	0.0	-0.8198	-0.2257	True
College	Senior Highschool	0.1424	0.5763	-0.1061	0.3909	False
Highschool	Inapplicable	-0.8751	0.0	-0.9405	-0.8097	True
Highschool	Middle School & Below	-0.003	1.0	-0.2794	0.2734	False
Highschool	Not-school	-0.8198	0.0	-1.0049	-0.6347	True
Highschool	Senior Highschool	-0.1547	0.0	-0.2428	-0.0666	True
Inapplicable	Middle School & Below	0.8721	0.0	0.5988	1.1454	True
Inapplicable	Not-school	0.0553	0.9529	-0.1252	0.2358	False
Inapplicable	Senior Highschool	0.7204	0.0	0.6424	0.7984	True
Middle School & Below	Not-school	-0.8168	0.0	-1.1404	-0.4932	True
Middle School & Below	Senior Highschool	-0.1516	0.6343	-0.4313	0.128	False
Not-school	Senior Highschool	0.6651	0.0	0.4752	0.8551	True

Table 3: Multiply Comparisons of Means for First-Sex Experience with p-value adjustments Tukey's Honestly Significant Difference (HSD) controlling Family-wise error rate (FWER)

Post-hoc comparisons were used to explore the pair-wise mean difference within groups of the variable through Tukey's HSD test indicated that several categories of first sex experience differed significantly in their mean scores which are the following:

- College scores were significantly higher than high school (mean difference = 0.30, p = .007).
- College scores were significantly lower than Inapplicable (mean difference = −0.58, p < .001).
- Highschool differed significantly from Inapplicable, Not-school, and Senior Highschool (all p < .001).
- Inapplicable had significantly higher scores than Middle School & Below and Senior Highschool.
- Not-school was significantly different from multiple groups, including College and Middle School & Below.

	sum_sq	df	F	PR(>F)
C(man_guardian)	25.044634	4.0	10.011252	4.537223e-08
Residual	3292.794852	5265.0	NaN	NaN

Table 4: One-Way ANOVA Test for Male Guardian

A one-way ANOVA was conducted to examine the effect of the presence of a male guardian on the dependent variable. There was a statistically significant effect of group, F(4, 5265) = 10.011252, p = 4.537223e-08.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Bio father	No father figure	0.0421	0.9301	-0.1012	0.1854	False
Bio father	Other father figure	0.0078	0.9999	-0.1357	0.1514	False
Bio father	Stepfather	0.1284	0.1628	-0.0275	0.2843	False
Bio father	With both parents	-0.1057	0.0184	-0.1996	-0.0117	True
No father figure	Other father figure	-0.0343	0.9781	-0.1956	0.127	False
No father figure	Stepfather	0.0863	0.6499	-0.0861	0.2587	False
No father figure	With both parents	-0.1478	0.0066	-0.2671	-0.0284	True
Other father figure	Stepfather	0.1205	0.3144	-0.0521	0.2931	False
Other father figure	With both parents	-0.1135	0.0726	-0.2332	0.0061	False
Stepfather	With both parents	-0.234	0.0	-0.3683	-0.0998	True

Table 5: Multiply Comparisons of Means for the presence of Male Guardian with p-value adjustments Tukey's Honestly Significant Difference (HSD) controlling Family-wise error rate (FWER)

Post-hoc comparisons were used to explore the pair-wise mean difference within groups of the variable through Tukey's HSD test indicating that there were only select categories of presence of Male Guardian that differed significantly in their mean scores which are the following:

- Respondents with both parents had significantly higher scores than the rest of the groups except for those groups that were identified to have another father figure.

	sum_sq	df	F	PR(>F)
C(woman_guardian)	22.097570	3.0	11.76112	1.122049e-07
Residual	3299.915355	5269.0	NaN	NaN

Table 6: One-Way ANOVA Test for Female Guardian

A one-way ANOVA was conducted to examine the effect of the presence of a female guardian on the dependent variable. There was a statistically significant effect of group, F(3, 5269) = 11.76112, p = 1.122049e-07.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Bio mother	No mother figure	0.0508	0.9618	-0.2165	0.3181	False
Bio mother	Other mother figure	0.0619	0.5194	-0.0543	0.1781	False
Bio mother	With both parents	-0.1185	0.0001	-0.1876	-0.0493	True
No mother figure	Other mother figure	0.0111	0.9996	-0.2676	0.2898	False
No mother figure	With both parents	-0.1692	0.3471	-0.4318	0.0933	False
Other mother figure	With both parents	-0.1803	0.0001	-0.2851	-0.0755	True

Table 7: Multiply Comparisons of Means for the presence of a Female Guardian with p-value adjustments Tukey's Honestly Significant Difference (HSD) controlling Family-wise error rate (FWER)

Post-hoc comparisons were used to explore the pair-wise mean difference within groups of the variable through Tukey's HSD test indicating that there were only select categories of the presence of a female guardian that differed significantly in their mean scores which are the following:

- Respondents with both parents had significantly higher scores than the rest of the groups except for those groups that were identified to have no mother figure.

	sum_sq	df	F	PR(>F)
C(marital_status)	230.736143	4.0	98.232524	9.466236e-81
Residual	3089.948046	5262.0	NaN	NaN

Table 8: One-Way ANOVA Test for Marital Status

A one-way ANOVA was conducted to examine the effect of the marital status of the respondents on the dependent variable. There was a statistically significant effect of group, F(4, 5262) = 98.232524, p = 9.466236e-81.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Divorced/annulled	Married	-0.5213	0.0	-0.6453	-0.3973	True
Divorced/annulled	Never married	-0.7653	0.0	-0.8866	-0.644	True
Divorced/annulled	Separated	-0.1756	0.2338	-0.4073	0.056	False
Divorced/annulled	Widowed	-0.2384	0.6192	-0.6987	0.2219	False
Married	Never married	-0.244	0.0	-0.3053	-0.1827	True
Married	Separated	0.3457	0.0001	0.139	0.5523	True
Married	Widowed	0.2829	0.4204	-0.1654	0.7312	False
Never married	Separated	0.5897	0.0	0.3847	0.7947	True
Never married	Widowed	0.5269	0.0116	0.0794	0.9745	True
Separated	Widowed	-0.0628	0.9968	-0.5519	0.4263	False

Table 9: Multiply Comparisons of Means for Marital Status with p-value adjustments Tukey's Honestly Significant Difference (HSD) controlling Family-wise error rate (FWER)

Post-hoc comparisons were used to explore the pair-wise mean difference within groups of the variable through Tukey's HSD test indicating that several categories of marital status differed significantly in their mean scores which are the following:

- Never-married groups have a higher mean score of sexual partners than those who were identified to be separated and widowed.
- Groups that are married have a lower mean score of sexual partners than those who were identified to have never married and separated.

- Groups that are divorced or annulled have a lower mean score of sexual partners than those who were identified to have never married and married.

To reduce the risk of false positives due to multiple pairwise comparisons, we used Tukey's Honestly Significant Difference (HSD) test which controls the family-wise error rate at 0.05, in order to ensure the robustness of the statistical test, ensuring that the overall probability of creating Type I errors are controlled, leading to more trustworthy conclusions and interpretations.

Poisson Regression

Variable Name	Description
OPPLIFENUM	Number of opposite-sex partners in lifetime for all types of sex (computed in FC J-14d) (top-coded)
Predictors:	
MARSTAT	AD-7b R's marital or cohabiting status
ONOWN	AG-0a (before age 18) R ever live away from parents/guardians?
LVSIT14F	AG-3 female parent (figure) living with at age 14 - fam not intact thru 18
LVSIT14M	AG-4 male parent (figure) living with at age 14 - fam not intact thru 18
MENARCHE	BA-1 Age at first menstrual period (bottom-coded)
GRFSTSX	CE-8 Grade R Was in at First Sexual Intercourse (bottom-coded)
WHOFSTPR	CG-3 Who Was Rs First Sexual Partner
RELIGION	Current religious affiliation
ECTIMESX	EA-12 Number of times R used emergency contraception
HIEDUC	Highest completed year of school or highest degree received (bottom-coded)
RELDLIFE	IC-7 How important is religion in R's daily life
DRINK12	JC-4 Last 12 mos: how often drank alcoholic beverages
POT12	JC-6 Last 12 mos: how often smoked marijuana
CONDSEXL	JD-11 Was condom used at last sex of any kind with a male partner
MALSHT12	JF-6 Last 12 mos: R had sex with male intravenous drug user
ABORTION	Number of completed pregnancies ending in induced abortion
CURRPRTT	Number of Current Male Sexual Partners-including curr H/P
PARTS1YR	Number of opposite-sex sexual partners in last 12 months (top-coded)
AGE_R	R's age at interview (FC A-2b)
RSCRRACE	R's race as reported in screener
TOTINCR	Total income of R's family

Table 10: A table containing the predictors and response variables (of female respondents) to be used in Poisson Regression Modeling.

2. Exploratory Data Analysis

2.1 Missing Data and Cardinality

	Counts of Null	nunique	unique
AGE_R	0	36	[29, 18, 37, 40, 49, 30, 25, 44, 28, 47, 21, 1...
RSCRRACE	0	4	[3, 4, 2, 1]
HIEDUC	0	11	[4, 5, 6, 10, 9, 8, 3, 1, 2, 7, 11]
RELIGION	0	4	[4, 2, 1, 3]
RELDLIFE	1559	5	[1.0, 2.0, nan, 3.0, 9.0, 8.0]
MARSTAT	0	5	[3, 2, 1, 8, 9]
vry1stag	836	34	[21.0, nan, 17.0, 16.0, 15.0, 22.0, 27.0, 14.0...
LVSIT14F	3098	5	[nan, 1.0, 8.0, 2.0, 3.0, 9.0]
LVSIT14M	3098	6	[nan, 3.0, 2.0, 1.0, 9.0, 8.0, 4.0]
MENARCHE	0	21	[9, 13, 12, 11, 14, 15, 10, 8, 16, 96, 18, 98,...
ECTIMESX	3184	17	[nan, 2.0, 5.0, 1.0, 4.0, 3.0, 10.0, 99.0, 8.0...
CONDSEXL	3546	4	[nan, 1.0, 5.0, 8.0, 9.0]
PARTS1YR	836	8	[0.0, nan, 1.0, 3.0, 2.0, 5.0, 4.0, 7.0, 6.0]
CURRPRTT	0	4	[0, 1, 2, 3]
DRINK12	19	8	[1.0, 3.0, 5.0, 4.0, 2.0, 6.0, 9.0, nan, 8.0]
MALSHT12	1243	4	[nan, 5.0, 1.0, 8.0, 9.0]
ABORTION	2053	7	[nan, 0.0, 1.0, 2.0, 6.0, 3.0, 5.0, 4.0]
WHOFSTPR	3290	8	[nan, 1.0, 7.0, 20.0, 9.0, 8.0, 98.0, 99.0, 2.0]
ONOWN	0	4	[5, 1, 8, 9]
POT12	22	8	[6.0, 1.0, 3.0, 4.0, 2.0, 5.0, 8.0, nan, 9.0]
GRFSTSX	2459	14	[nan, 10.0, 11.0, 9.0, 96.0, 7.0, 15.0, 8.0, 1...
TOTINCR	0	15	[15, 11, 13, 14, 7, 8, 1, 6, 12, 4, 2, 9, 10, ...
OPPLIFENUM	761	47	[1.0, nan, 9.0, 8.0, 7.0, 4.0, 15.0, 13.0, 5.0...

Table 11: A table representing variable characteristics for sexual behavior analysis of female respondents

Based on Table 11, the data set of *Female Respondents* contains many null values. Null values are placeholders for the respondents who are not applicable to the specific questions. There are significant number of instances where a respondent “Don’t Know” (9,99,999) or “Refused” (8,88,88) to answer the question.

Each variable required different handling since some of the seemingly unusable values may contain key information.

2.2 Distributions

2.2.1 Continuous Variable

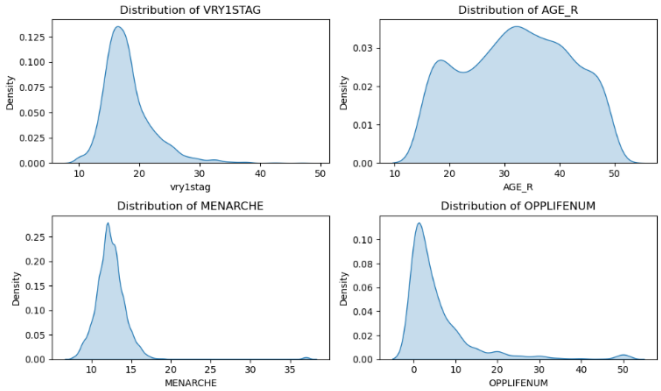


Figure 6: Distribution of Continuous Variables (with cardinality of 21 and above)

Without the invalid values (“Don’t Knows” and “Refused”), Figure 1 shows the distribution of continuous variables. The dependent variable “OPPLIFENUM,” is skewed to the right which is to be expected of a count variable. The same is the case for “Menarche” and “VRY1STAG.” “AGE_R” appeared to be relatively normally distributed (compared to others).

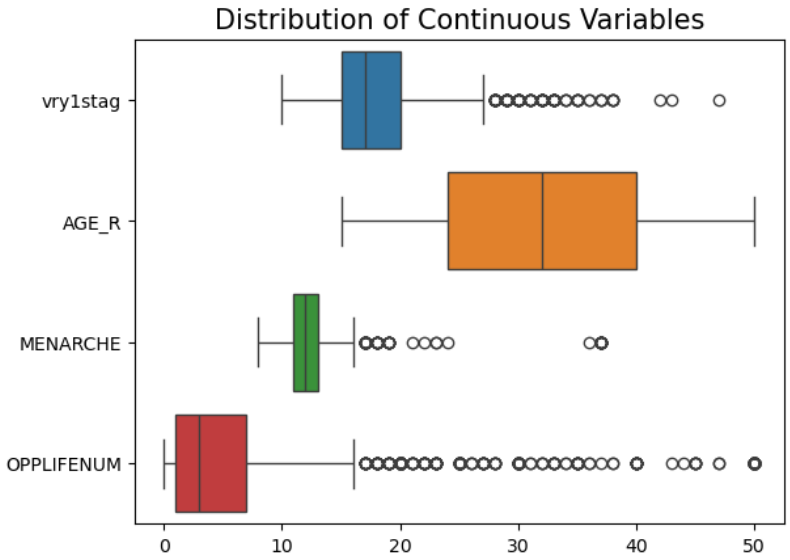


Figure 7: Box-Plot Distribution of Continuous Variables (with cardinality of 21 and above)

Figure 7 showed the Box-Plot distribution of the continuous variables and the presence of some significant outliers. A normal regression will not be suitable for this dataset for even the "Opplifenum" have numerous significant outliers. Since that is normal behavior for count variables, these outliers could not be removed. *Poisson Regression* was used instead of normal regression. Although the presence of outliers may be an early warning of *overdispersion*. This will be problematic with *Poisson Regression* because the primary assumption of it is that the *mean is equal to variance*. Overdispersion occurs when this assumption is violated. In such cases, *Negative Binomial Regression* will be used.

2.2.1 Categorical Variable

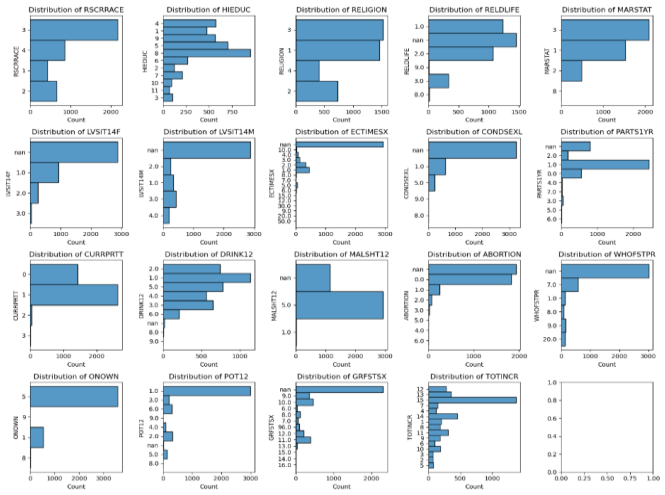


Figure 8: Distribution of Discrete/Categorical Variable. Note that the histogram is horizontally oriented.

Figure 3 indicates a highly imbalanced distribution of categorical variables. Some features have significant “nan” or respondents whose inapplicable for a specific question, sometimes even greater than those who are applicable. This is an indication that one-hot-encoding is necessary to counteract the imbalance.

2.3 Variable Transformation Details

Variable	Wrangling Steps
ABORTION	• NaN values filled with 0. • Renamed to Abortion_num.
CONDSEXL	• NaN values replaced with "inapplicable". • Value 5 replaced with 0.<p> • One-Hot Encoded.
DRINK12	• One-Hot Encoded. • Rows where DRINK12 is 9 are dropped.
ECTIMESX	• NaN values replaced with "Never" • Non-"Never" values are clustered using a KMedoids clustering model • Original values are replaced with cluster labels.
GRFSTSX	• NaN values filled with "nosex_u18 • Rows where GRFSTSX is 98 or 99 are dropped • Values 14, 15, and 16 are mapped to 13.
LVSIT14F	• Rows where LVSIT14F is 8 or 9 are dropped. • Values mapped using dictionary: {1: "Bio/adoptive mother", 2: "Other mother figure", 3: "No mother figure", None: "both parents"}. • One-Hot Encoded. • The column LVSIT14F_both parents is dropped.
LVSIT14M	• Rows where LVSIT14M is 8 or 9 are dropped. • Values mapped using dictionary: {1: "Bio/adoptive mother", 2: "Bio/adoptive father", 3: "Step father", 4: "Other father figure", None: "both parents"}. • One-Hot Encoded.
MALSHT12	• Each unique value transformed into separate binary indicator column.<p>• Original column dropped. • Rows where MALSHT12 is 8 or 9 are dropped.
MARSTAT	• Rows where MARSTAT is 8 or 9 are dropped.
MENARCHE	• Value 96 replaced with 37 • Rows where MENARCHE is 98 or 99 are dropped.
ONOWN	• Rows where ONOWN is 8 or 9 are dropped.
OPPLIFENUM	• NaN values filled with 0 • Rows where OPPLIFENUM is 998 or 999 are dropped (Invalid "Refused and Don't Know" responses).
PARTS1YR	• Value 5 replaced with 0 • Transformed into two binary columns: PARTS1YR_no and PARTS1YR_yes. • Original column dropped.
POT12	• One-Hot Encoded. • Rows where POT12 is 9 are dropped.
RELDLIFE	• Values replaced: 8 → 4, 9 → 5, NaN → 6.
WHOFSTPR	• Each unique value transformed into separate binary indicator column.<p>• Original column dropped. • Rows where WHOFSTPR is 98 or 99 are dropped.
vry1stag	• Rows where vry1stag is 97 are dropped. • Binned into categories: '≤14 years', '15–17 years', '18–19 years', '20+ years'.<p> • NaN values become "Never Had sex". • Binned version is One-Hot Encoded.

Table 12: A table representing the processing done to each variable. Code in appendix.

2.4 Correlation

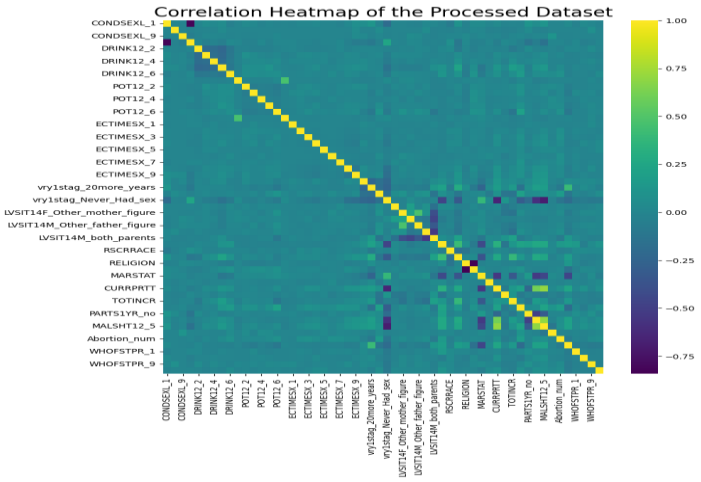


Figure 9: Correlational heatmap of the processed dataset. Note that the highest Pearson correlation is "CURRPRTT" and "MALSHT12_5" with 0.70. This correlation is not significant enough to flag multicollinearity.

3. Split

The original data were loaded again and was split into two sets, training set (80%) and test set (20%). The reloading of original state was done to prevent data leakage.

4. Cross Validation

A cross validation was conducted for each combination of parameters:

For Poisson Regression, the parameters "alpha" and "L1_wt" of regularization was tuned with the following set of values:

- alphas = [0.01, 0.1, 0.5, 1.0, 10.0]
- l1_wts = [0.1, 0.5, 0.9, 1.0]

For Negative Binomial Regression, "alpha_binom" was added to tune the best alpha for the Negative Binomial Regression object.

- alphas_binom_to_tune = [0.1, 0.5, 1.0, 2.0]

Note that calculation of deviance for a regularized model ("elastic_net" was used) is not yet implemented in python, or any statistical summary for that matter. Two probable reason for this are: (a) the fact that there is no consensus to the right computation some statistical summary of a regularized model (including deviance) and (b) the regularization fitting is primarily used for model selection that relies more heavily on cross-validated predictive performance rather than statistical metric performance.

To get around calculation of overdispersion ("overdis_params"), the modeler took the non-zero coefficient parameters of the regularized model and train a unregularized model to calculate for possible overdispersion.

Poisson Regression

	alpha	L1_wt	avg_mse	overdis_params
0	0.01	0.1	54.226508	4.290659
1	0.01	0.5	54.334660	4.296870
3	0.01	1.0	54.370134	4.296870
2	0.01	0.9	54.463904	4.320549
4	0.10	0.1	55.721707	4.333666
5	0.10	0.5	57.206010	4.372102
6	0.10	0.9	58.202503	4.697114
7	0.10	1.0	58.267722	4.737102
8	0.50	0.1	59.902721	4.419589
12	1.00	0.1	63.493816	4.806298

Table 13.1: Result of hyperparameter tuning of Poisson Regression (top 10 Mean Square Error or MSE)

Negative Binomial Regression

	alpha	L1_wt	alpha_binom	avg_mse	overdisp
3	0.01	1.0	0.1	55.022959	2.265840
1	0.01	0.5	0.1	55.259540	2.238389
2	0.01	0.9	0.1	55.375307	2.262001
0	0.01	0.1	0.1	55.647682	2.239168
17	0.01	0.5	0.5	57.160240	0.966760
16	0.01	0.1	0.5	57.248578	0.952015
18	0.01	0.9	0.5	57.279296	0.971139
4	0.10	0.1	0.1	57.411768	2.279693
19	0.01	1.0	0.5	57.706256	0.975821
33	0.01	0.5	1.0	57.920992	0.612056

Table 13.2: Result of hyperparameter tuning of Negative Binomial Regression (top 10 MSE)

Both Table 13.1 and 13.2 showed of the hyperparameter tuning and cross-validation for Poisson and Negative Binomial Regression, respectively. The Poisson Regression did have the higher mean square error. Although, as was suspected earlier (see Figure 7), overdispersion occur in all of them. For that reason, Negative Binomial Regression was imperative. Negative Binomial Regression did have some model that have over dispersed.

5. Model Selection

- Poisson Regression – {“alpha”: 0.01, “L1_wt”: 0.1} or the highest MSE (overdispersion =4.29).
- Negative Binomial Regression – {“alpha”: 0.01, “L1_wt”:0.5, “alpha_binom”: 0.5} or the highest MSE that does not over dispersed (overdispersion =0.97).

6. Model Evaluation

	model	params	test_mse	train_mse
0	Baseline	0	80.356538	79.021221
1	Binom_unr	33	56.258788	56.628048
2	Binom_reg	56	55.362855	55.132599
3	Pois_unr	43	54.054384	51.617108
4	Pois_reg	56	54.461459	53.612255

Table 14: MSE Result of each model. Note that the baseline is basically using the mean of the response variable as prediction for all values of predictors.

Table 14 clearly indicates that both Poisson Regression and Negative Binomial model performed significantly better than the baseline. Predictive accuracy speaking, it is ideal to pick the model with lowest MSE. In this case, Poisson Regression. But as was said earlier, all Poisson Regression model have resulted to overdispersion, including the best parameter. Thus, it becomes imperative to choose the **Negative Binomial model even though it performed worse since it didn't over dispersed.**

7. Communication

Model:	GLM	AIC:	19423.2079
Link Function:	Log	BIC:	-29858.7939
Dependent Variable:	OPPLIFENUM	Log-Likelihood:	-9678.6
Date:	2025-05-18 10:26	LL-Null:	-12650.
No. Observations:	4095	Deviance:	3927.0
Df Model:	32	Pearson chi2:	1.17e+04
Df Residuals:	4062	Scale:	1.0000
Method:	IRLS		

Table 15: Summary of unregularized Negative Binomial Regression trained using parameters from regularized Negative Binomial Regression

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.3088	0.1050	2.9413	0.0033	0.1030	0.5145
CONDSEXL_1	-0.0906	0.0390	-2.3247	0.0201	-0.1670	-0.0142
DRINK12_3	0.0128	0.0406	0.3144	0.7532	-0.0668	0.0923
DRINK12_5	0.1252	0.0380	3.2961	0.0010	0.0508	0.1997
DRINK12_6	0.2387	0.0607	3.9304	0.0001	0.1197	0.3578
POT12_2	0.3163	0.0516	6.1345	0.0000	0.2153	0.4174
POT12_3	0.2650	0.0612	4.3309	0.0000	0.1451	0.3849
POT12_4	0.7481	0.0852	8.7859	0.0000	0.5813	0.9150
POT12_5	0.3603	0.0735	4.9016	0.0000	0.2162	0.5044
POT12_6	0.5793	0.0502	11.5381	0.0000	0.4809	0.6777
ECTIMESX_3	0.3498	0.0948	3.6889	0.0002	0.1639	0.5356
ECTIMESX_4	0.2502	0.0716	3.4962	0.0005	0.1100	0.3905
ECTIMESX_7	0.4998	0.1213	4.1198	0.0000	0.2620	0.7375
ECTIMESX_8	0.1823	0.0483	3.7769	0.0002	0.0877	0.2769
ECTIMESX_9	0.1434	0.0430	3.3309	0.0009	0.0590	0.2277
vry1stag_18_19_years	-0.2310	0.0399	-5.7859	0.0000	-0.3093	-0.1528
vry1stag_20more_years	-0.5654	0.0435	-13.0124	0.0000	-0.6506	-0.4802
vry1stag_leq14_years	0.2059	0.0442	4.6600	0.0000	0.1193	0.2925
vry1stag_Never_Had_sex	-2.2914	0.0791	-28.9551	0.0000	-2.4465	-2.1363
LVSIT14M_Step_father	0.1142	0.0451	2.5326	0.0113	0.0258	0.2027
AGE_R	0.0281	0.0019	15.0298	0.0000	0.0245	0.0318
HIEDUC	0.0652	0.0067	9.7670	0.0000	0.0521	0.0783
RELIGION	-0.0282	0.0140	-2.0108	0.0443	-0.0557	-0.0007
ONOWN	-0.0122	0.0102	-1.1943	0.2323	-0.0323	0.0078
PARTS1YR_yes	-0.6433	0.0377	-17.0618	0.0000	-0.7172	-0.5694
MALSHT12_5	0.9277	0.0512	18.1261	0.0000	0.8274	1.0280
MALSHT12_1	1.3465	0.1765	7.6283	0.0000	1.0005	1.6924
Abortion_num	0.0787	0.0268	2.9328	0.0034	0.0261	0.1313
WHOFSTPR_7	-1.3432	0.0530	-25.3215	0.0000	-1.4472	-1.2392
WHOFSTPR_1	-0.2699	0.0736	-3.6653	0.0002	-0.4142	-0.1256
WHOFSTPR_8	-1.1815	0.1089	-10.8470	0.0000	-1.3949	-0.9680
WHOFSTPR_9	-0.1890	0.0697	-2.7128	0.0067	-0.3255	-0.0524
WHOFSTPR_20	0.0795	0.0705	1.1270	0.2598	-0.0588	0.2178

Table 16: Coefficients of unregularized Negative Binomial Regression trained using parameters from regularized Negative Binomial Regression

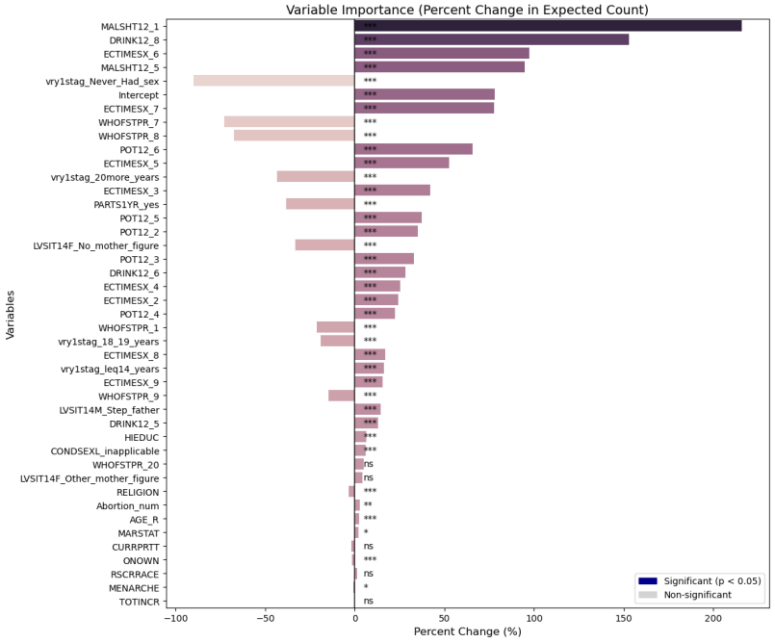


Figure 10: Variable Importance for selected parameters of Negative Binomial Regression. Note that Percent Change (%) was calculated by

$100 * e^{\beta_k - 1} \quad \forall k \in params_{reg}$ where β_k is the coeficient of parameter k

Figure 10 shows how different factors affect the the final model, measured as percent change in expected count. Most variables in the model are statistically significant (marked with ***).

Strong Positive Effects:

- “MALSHT12_1” has the largest impact, increasing expected counts by ~200%. These are the people who had sex with male drug user in the past 12 months.

- “DRINK12_8” increases counts by ~150%. These are the people who have refused to disclosed how often they drank alcohol in the past 12 months.

- ECTIMESX_6 both increase counts by ~100%. These are the cluster of people who have used emergency contraceptives about 15-20 times.

Strong Negative Effects:

- “vry1stag_Never_Had_sex” decreases counts by ~75%. These are the people who never had sex.

- “WHOFSTPR” variables (7 and 8) decrease counts by ~50%. WHOFSTPR_7 and WHOFSTPR_8 are the people who have had their first sexual encounter with their husband or current partner.

- “vry1stag_20more_years” decreases counts by ~50%. These are the people who has had their sexual encounter when their 20 and older.

Contingency Tables

Variable Name	Description
AGE_R	Age of respondent
HIEDUC	Highest educational attainment
POVERTY	Family income as percentage of federal poverty threshold
RELIGION	Current religious affiliation
INTCTFAM	Intact status of childhood family
vry1stag	Age at first sexual intercourse
FMARITAL	Marital Status
CONSTAT1	Current contraceptive status
GENHEALT	General health status (scaled from Poor to Excellent)
OPPLIFENUM	Number of opposite-sex partners in lifetime

Table 17: Variables Used in Contingency Tables Section

1. Education vs. Contraceptive Use

HIEDUC	Below 9	9–11	High school/GED	Some college	Associate	Bachelor	Master	Doctorate
CONSTAT1_GROUPED								
Barrier	17	11	2	46	55	26	13	99
Hormonal	92	20	15	141	181	42	36	228
IUD	19	6	11	39	65	33	30	152
Natural	9	6	10	42	65	22	14	83

Nonuser	349	61	36	274	269	79	77	309
Other	29	17	16	81	69	26	25	110
Sterile	103	45	62	169	241	121	82	274

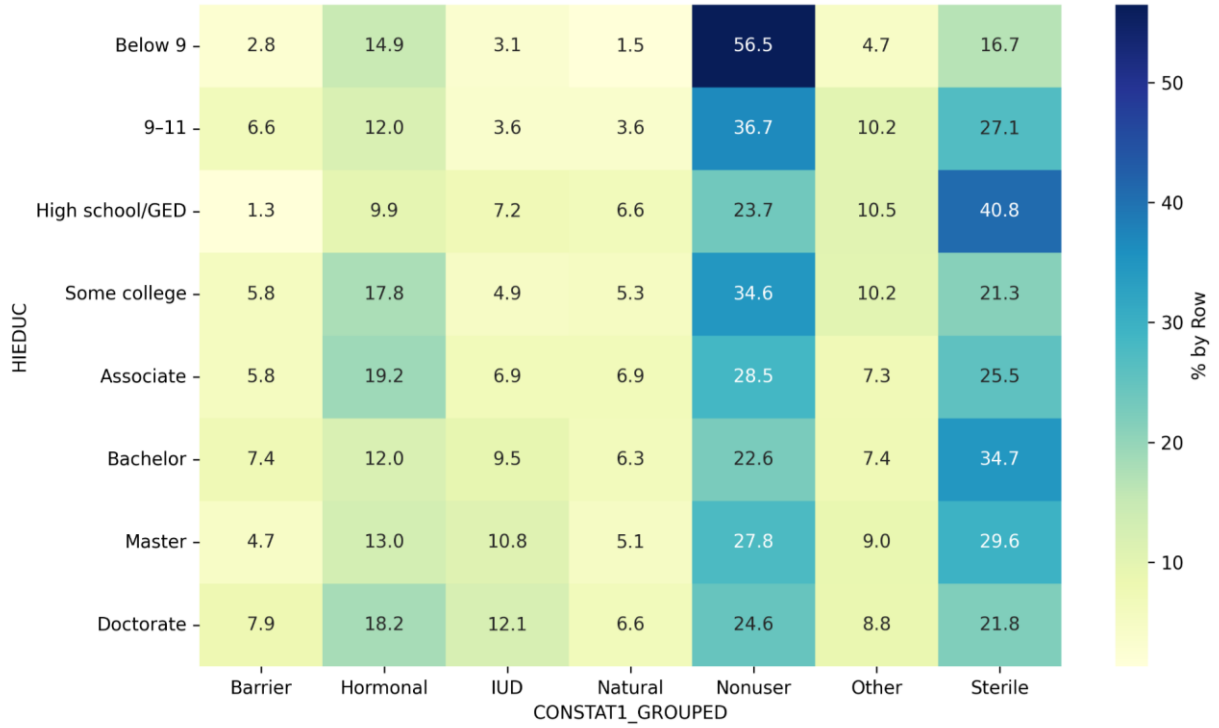


Figure 11. Education vs. Contraceptive Use (Row-wise %)

2. Religion vs. Age at First Sexual Intercourse

vry1stag	Under 13	13-17	18-23	Above 23
RELIGION				
Catholic	18	289	251	259
No Religion	46	736	406	380

Others	16	138	104	154
Protestant	61	811	492	393

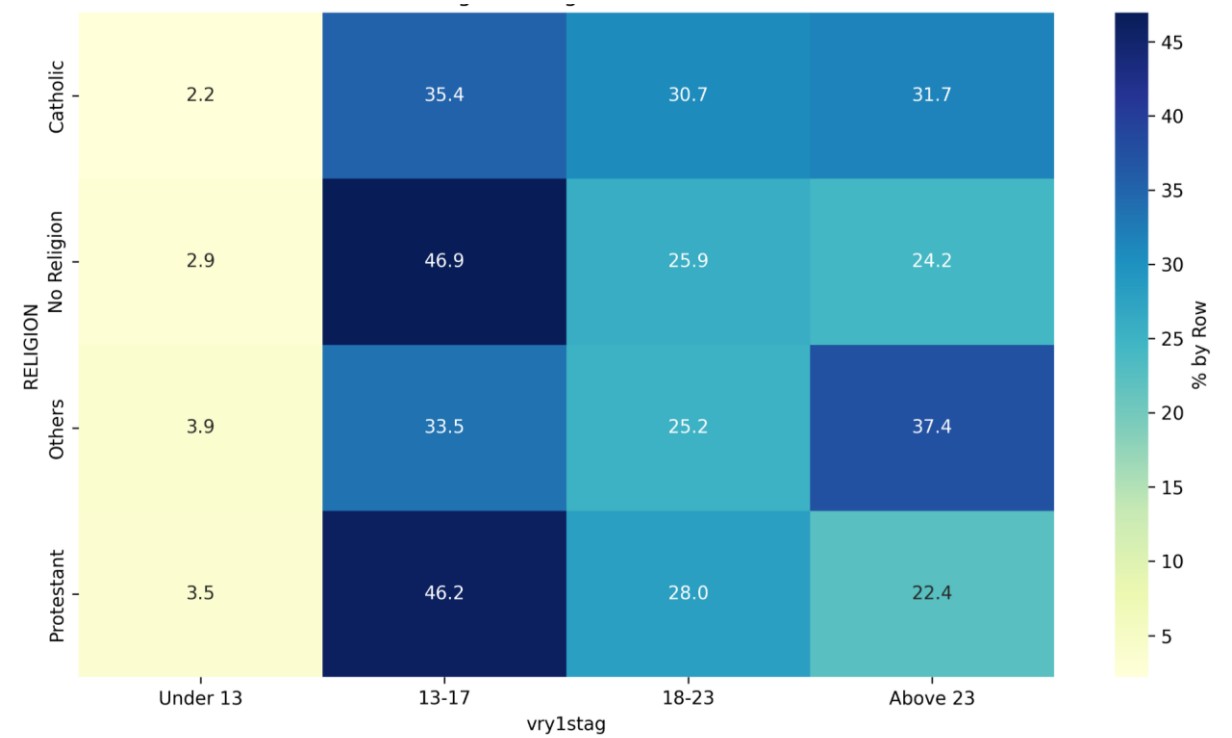


Figure 12. Religion vs. Age at First Sexual Intercourse (Row-wise %)

3. General Health vs. Number of Opposite-Sex Partners

OPPLIFENUM	One	Few	Several	Many	Plenty

GENHEALT					
Excellent	197	200	153	82	297
Very Good	285	370	275	162	483
Good	227	334	299	173	455
Fair	61	98	91	47	162
Poor	12	19	24	14	34

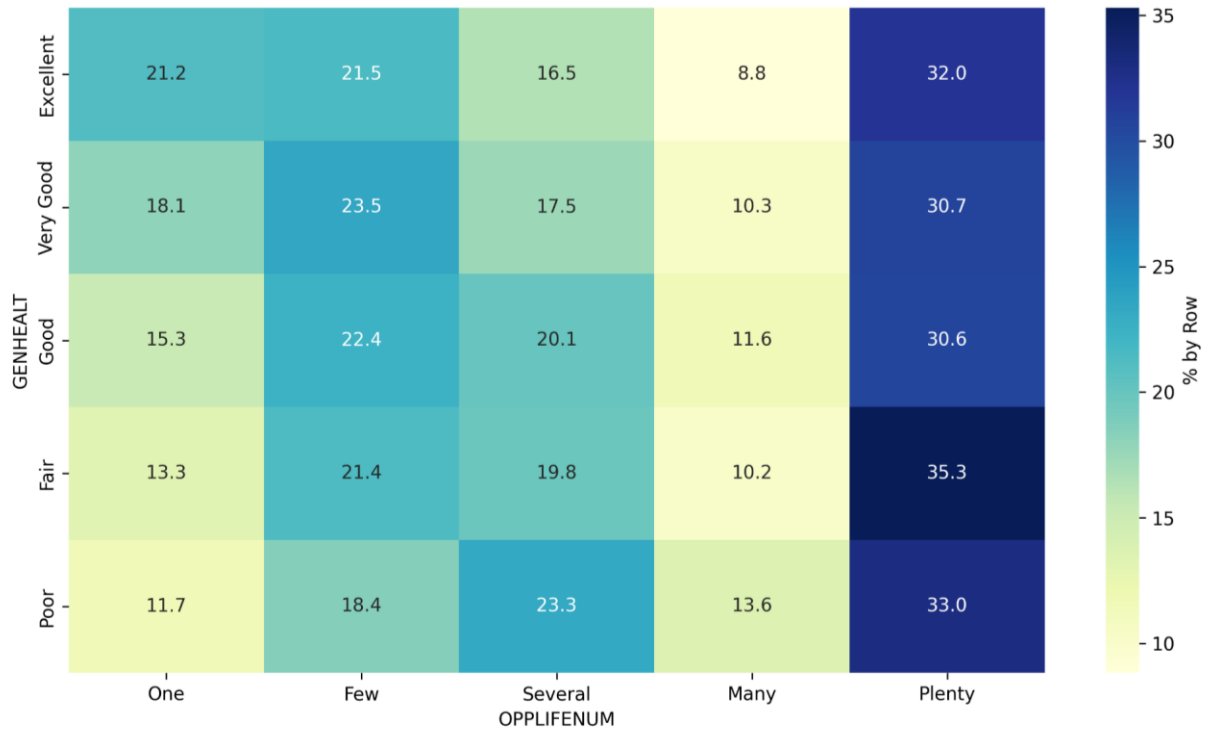


Figure 13. General Health vs. Number of Opposite-Sex Partners (Row-wise %)

4. Poverty Level vs. Marital Status

FMARITAL POVERTY	Divorced	Married	Never married	Separated	Widowed
Below Poverty	83	157	737	48	7
Near Poverty	50	263	609	28	7
Above Poverty	97	440	669	21	3
Well Above Poverty	66	593	665	9	2

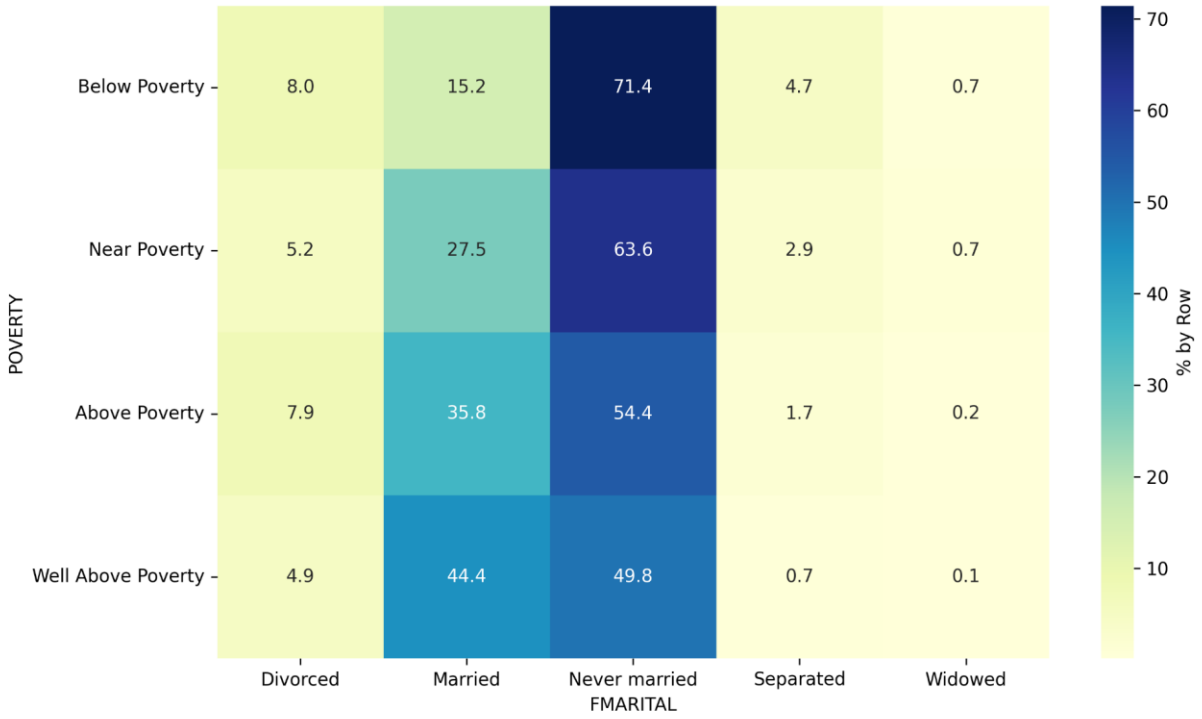


Figure 14. Poverty Level vs. Marital Status (Row-wise %)

Chi-Square Test

Comparison	Chi-square	p-value	df
Education vs. Contraceptive Use	371.582669	2.266653e-54	42
Religion vs. Age at First Sexual Intercourse	79.875308	1.710974e-13	9
General health vs. Number of opposite-sex partner	35.777128	3.106796e-03	16
Poverty level vs. marital status	286.406587	3.338153e-54	12

1. **Education vs. Contraceptive Use**

   There is a high and statistically significant association between education level and contraceptive method used. The very low p-value indicates that contraceptive preferences vary significantly across educational backgrounds. This makes sense—education often affects knowledge, access, and attitudes toward reproductive health.

2. **Religion vs. Age at First Sex**

   There is a strong and significant association between religious affiliation and the age at first sexual intercourse. Religious beliefs may influence norms around abstinence, sexual behavior, and the timing of first sex. The relationship is unlikely to be due to chance.

3. **General Health vs. Number of Opposite-Sex Partners**

   There is a statistically significant association, though weaker than others. General health status may be related to the number of opposite-sex partners—possibly reflecting behavioral or socioeconomic factors affecting both.

4. **Poverty Level vs. Marital Status**

   There is an extremely strong and significant association between poverty level and marital status. Lower-income individuals may have different patterns in marriage or separation compared to those above the poverty line—possibly due to economic barriers, social norms, or stress-related factors.

Categorical Response Modeling

The variables used are the same as the previous section, “Contingency Tables.”

Ordinal Logistic Regression Analysis

Using ordinal logistic regression, we will examine the relationship between age, education, income level, religion, childhood family status, age at first sexual intercourse, marital status, contraceptive use, general health, and the likelihood of having multiple opposite-sex partners. We will use Python libraries, such as statsmodels, to apply the regression model.

1. Model Fitting

Metric	Value
Log-Likelihood	-6696.79871
AIC	13439.597419
Converged	True

a. Log-Likelihood

The model has a log-likelihood of -6696.8, which indicates an acceptable fit for ordinal logistic regression on social data.

b. AIC (Akaike Information Criterion)

The AIC value of 13,440 suggests a balance between model fit and simplicity. This can be used for comparison if multiple models are tested.

c. Convergence

The model converged successfully, indicating the estimated parameters are statistically reliable.

2. Odds Ratios

	Coefficient	Odds Ratio	CI Lower (95%)	CI Upper (95%)	p-value
AGE_R	-0.16511	0.8478	0.783705	0.917138	3.85E-05
HIEDUC	-0.028148	0.972245	0.945658	0.999578	4.66E-02

POVERTY	0.086903	1.090791	1.034576	1.150062	1.29E-03
vry1stag	0.211503	1.235534	1.149587	1.327907	8.95E-09
GENHEALT	0.099107	1.104184	1.0452	1.166497	4.03E-04
RELIGION_No Religion	0.507521	1.661168	1.417341	1.946941	3.69E-10
RELIGION_Others	0.154429	1.166992	0.93359	1.458746	1.75E-01
RELIGION_Protestant	0.349327	1.418113	1.215232	1.654865	9.22E-06
INTCTFAM_Had two parents	-0.250334	0.778541	0.692835	0.874849	2.59E-05
FMARITAL_Married	-1.019873	0.360641	0.290479	0.44775	2.48E-20
FMARITAL_Never married	-0.2419	0.785134	0.632387	0.974777	2.84E-02
FMARITAL_Separated	-0.457586	0.63281	0.4351	0.920359	1.67E-02
FMARITAL_Widowed	-0.631696	0.531689	0.232401	1.216404	1.35E-01
CONSTAT1_GROUPED_Hormonal	0.347272	1.415201	1.10535	1.81191	5.88E-03
CONSTAT1_GROUPED_IUD	0.710979	2.035983	1.53973	2.692177	6.10E-07
CONSTAT1_GROUPED_Natural	0.576414	1.779645	1.320849	2.397804	1.51E-04
CONSTAT1_GROUPED_Nonuser	1.217933	3.380193	2.660314	4.294871	2.11E-23
CONSTAT1_GROUPED_Other	0.512494	1.66945	1.265039	2.203144	2.93E-04
CONSTAT1_GROUPED_Sterile	0.792502	2.208915	1.73735	2.808477	9.92E-11

a. **AGE_R** (Age)
    i. For each increase in age group, the odds of being in a higher category of opposite-sex partners decrease by 15.2%.
    ii. Older respondents are less likely to report more partners.
b. **HIEDUC** (Education Level)
    i. Not statistically significant at 0.05 (p = 0.047 borderline).
    ii. Slight negative effect: more education may be associated with fewer partners, but this effect is weak.
c. **POVERTY** (Poverty Level)
    i. For each increase in poverty level, the odds of reporting more partners increase by 9.1%.
    ii. Suggests possible link between economic status and sexual behavior.
d. **vry1stag** (Age at First Sex)

i. Each increase in age group of first sex increases odds of more partners by 23.6%.

ii. This seems counterintuitive — may reflect categorization. Possibly, "older first sex" may cluster with specific behaviors later.

e. **GENHEALT** (General Health)

Better general health is associated with 10.4% higher odds of reporting more partners.

f. **RELIGION**

i. **No Religion**: 66% higher odds of having more partners than Catholics (baseline).

ii. **Protestant**: 41.8% higher odds.

g. **INTCTFAM_Had two parents** (Had two parents)

Individuals from two-parent households are 22.2% less likely to report higher partner counts.

h. **FMARITAL**** (Marital Status)

i. **Married**: 64% lower odds of more partners.

ii. **Never married**: 33% lower odds

iii. **Separated/Widowed**: Even lower odds, suggesting marital status strongly predicts behavior.

i. **CONSTAT1_GROUPED**

i. **Hormonal**: Users are 68% more likely to be in higher categories

ii. **IUD**: Double the odds of reporting more partners

iii. **Natural**: Strong positive association

iv. **Nonuser**: Very strong predictor of high partner count

v. **Other**: Similar to hormonal methods

vi. **Sterile**: Strong positive odds

Overall, the ordinal logistic regression model significantly predicts the number of opposite-sex partners based on variables such as age, poverty level, marital status,

religion, and contraceptive method. For example, individuals using IUDs had 2.0 times higher odds of reporting more partners compared to the baseline group. Nonusers had 3.4 times higher odds. Additionally, being older, married, or from a two-parent household decreased the likelihood of reporting higher partner counts. The model converged successfully, and fit statistics (AIC = 13,440) indicate acceptable performance for social science data.

Multinomial Logistic Regression Analysis

Using multinomial logistic regression, we will examine the relationship between age, education, income level, religion, childhood family status, age at first sexual intercourse, marital status, number of sex partners had, general health, and the likelihood of using a specific contraceptive method. We will use Python libraries again, such as statsmodels, to apply the regression model.

1. **Model Fitting**

Metric	Value
Log-Likelihood	-6696.79871
AIC	13439.597419
Converged	True

The multinomial logistic regression model converged successfully, indicating that the optimization algorithm reached a stable solution and that the estimated coefficients are statistically valid. The model produced a log-likelihood of -6858.41, which reflects the degree of model fit—the higher (i.e., less negative) the value, the better the model fits the observed data.

The Akaike Information Criterion (AIC) value of 13,896.82 suggests an acceptable balance between model fit and complexity. Although AIC is not interpretable in absolute terms, it serves as a key comparative metric; lower AIC values are generally preferred when comparing alternative models.

2. **Odds Ratios**

A positive coefficient indicates increased odds of choosing that method over the base category, while a negative coefficient means decreased odds. Below are variables with their respective metrics. Only statistically significant predictors ($p < 0.05$) are highlighted.

Variable	Outcome	Coefficient	Odds Ratio	p-value
AGE_R	Hormonal	-0.2363	0.7895437753	0.026
POVERTY	Hormonal	0.2443	1.276727291	0.001
OPPLIFENUM	Hormonal	0.1913	1.210822644	0.001
AGE_R	IUD	0.2818	1.325513591	0.023
POVERTY	IUD	0.1723	1.18803419	0.039
GENHEALT	IUD	0.2168	1.242095658	0.013
OPPLIFENUM	IUD	0.2702	1.31022647	0
vry1stag	Natural	-0.4037	0.6678444446	0.001
HIEDUC	Nonuser	-0.1314	0.8768669574	0
vry1stag	Nonuser	0.7372	2.090075103	0
OPPLIFENUM	Nonuser	0.402	1.494811333	0
HIEDUC	Other	-0.0873	0.9164021338	0.042
OPPLIFENUM	Other	0.2239	1.250945919	0
AGE_R	Sterile	1.0026	2.725358557	0
HIEDUC	Sterile	-0.1731	0.8410535053	0
GENHEALT	Sterile	0.2816	1.325248515	0
OPPLIFENUM	Sterile	0.2939	1.341649732	0

a. Hormonal Method (vs. Base)

i. **AGE_R**: Coef = −0.2363, OR ≈ 0.79
Older respondents are **21% less likely** to choose hormonal methods. *(p = 0.026)*

ii. **POVERTY**: Coef = 0.2443, OR ≈ 1.28
Higher poverty level is associated with a **28% increase** in odds of choosing hormonal methods. *(p = 0.001)*

iii. **OPPLIFENUM**: Coef = 0.1913, OR ≈ 1.21
Those with more opposite-sex partners have **21% higher odds** of choosing hormonal methods. *(p < 0.001)*

b. IUD

i. **AGE_R**: Coef = 0.2818, OR ≈ 1.33
Older respondents are **33% more likely** to choose IUD. *(p = 0.023)*

ii. **POVERTY**: Coef = 0.1723, OR ≈ 1.19
More poverty → higher odds of IUD use. *(p = 0.039)*

iii. **GENHEALT**: Coef = 0.2168, OR ≈ 1.24
Better health associated with IUD use. *(p = 0.013)*

iv. **OPPLIFENUM**: Coef = 0.2702, OR ≈ 1.31
Strong predictor: more partners → **31% higher odds** of using IUD. *(p < 0.001)*

c. Natural Methods

i. **vry1stag**: Coef = −0.4037, OR ≈ 0.67
→ Later age at first sex **reduces odds** of using natural methods by **33%**. *(p = 0.001)*

d. Nonuser

i. **HIEDUC**: Coef = −0.1314, OR ≈ 0.88
More educated respondents are **12% less likely** to be nonusers. *(p < 0.001)*

ii. **vry1stag**: Coef = 0.7372, OR ≈ 2.09
Later sexual debut strongly increases likelihood of being a nonuser. *(p < 0.001)*

iii. **OPPLIFENUM**: Coef = 0.4020, OR ≈ 1.50
More sexual partners = **50% more likely** to be nonusers. *(p < 0.001)*

e. Other Methods

i. **HIEDUC**: Coef = −0.0873, OR ≈ 0.92
Higher education **reduces odds** of choosing other methods by **8%**. *(p = 0.042)*

ii. **OPPLIFENUM**: Coef = 0.2239, OR ≈ 1.25
More partners = **25% more likely** to choose "Other". *(p < 0.001)*

f. Sterile

i. **AGE_R**: Coef = 1.0026, OR ≈ 2.72
Older respondents are nearly **3x more likely** to use sterilization. *(p < 0.001)*

ii. **HIEDUC**: Coef = −0.1731, OR ≈ 0.84
More educated = **16% less likely** to choose sterilization. *(p < 0.001)*

iii. **GENHEALT**: Coef = 0.2816, OR ≈ 1.33
Better health = **33% higher odds** of sterilization. *(p < 0.001)*

iv. **OPPLIFENUM**: Coef = 0.2939, OR ≈ 1.34
Strong positive predictor. *(p < 0.001)*

Overall, the results of the multinomial logistic regression suggest that several demographic and behavioral factors significantly influence contraceptive method choice. For example, individuals with a higher number of opposite-sex partners were consistently more likely to choose IUDs, hormonal methods, and sterilization. Conversely, higher educational attainment was associated with a decreased likelihood of being a nonuser or choosing sterilization.

Age, poverty level, general health, and age at first sexual intercourse also emerged as significant predictors in multiple contraceptive categories. Notably, increasing age was strongly associated with higher odds of selecting long-term or permanent methods such as IUDs and sterilization, while later age at first sex was linked to reduced reliance on natural methods and increased likelihood of nonuse.

Conclusion

The analysis of female respondents from the NSFG dataset successfully identified several key factors influencing reproductive health behaviors.

Predictors such as age, education level, poverty status, age at first sexual intercourse, religious affiliation, family structure during childhood, and current contraceptive use are some of the more significant predictors in predicting the number of opposite-sex partners of female respondents. It is also worth noting that behaviors such as prior sex with a male drug user and high emergency contraceptive use were associated with a higher number of sexual partners, while never having had sex or having first sexual experiences later in life or with a husband/current partner were linked to fewer partners.

With regards to contraceptive choice, demographic and behavioral factors like age, poverty level, general health, age at first sexual intercourse, and the number of opposite-sex partners significantly influenced the type of contraceptive method used. For instance,

individuals with more partners were more likely to use IUDs, hormonal methods, or sterilization. Higher educational attainment was also generally associated with a decreased likelihood of being a nonuser or choosing sterilization.

Recommendations

It is important to **develop tailored educational programs and resources** focusing on groups identified with higher-risk behaviors or specific contraceptive needs (e.g., younger individuals, those with earlier sexual debut, those with a higher number of partners).

Given the influence of age at first sex, **further explore how to best support informed decision-making around sexual debut.**

Enhance comprehensive **sexual health education**, emphasizing various contraceptive methods, their effectiveness, and addressing socioeconomic barriers that may limit access to preferred options, particularly considering the influence of poverty.

Explore the **underlying reasons for the observed associations**, particularly with the strong links between specific behaviors (e.g., emergency contraceptive use frequency, substance use in partners) and the number of sexual partners.

Investigate the dynamics influencing contraceptive choices in more detail, like looking into the male perspective (dataset) and cross-analyzing with the female perspective. Even just digging deeper to the objective of this report like how life transitions and changing partner numbers affect method selection over time.

Code Appendix

The annotated codes or notebook file, divided by sections, can be found here:

https://github.com/RomandRapido/DSC1105_EDA/blob/main/SA2_Ramilo-Lansangan-Dacanay/Delivarables_CODES-SEC%201-SA2%20GROUP%201-DACANAY%2C%20J%3B%20LANSANGAN%2C%20R%3B%20RAMILO%2C%20Z.ipynb