Is Gender a Significant Factor in the Increase of the Number of Heterosexual Partners in the Last Decade?

Group A.1: Allan, Bhargav, Stefanie

The survey

The survey chosen for this research question is a survey conducted between 1999 and 2001 and between 2010 and 2012 under the name NSSAL-2 and NSSAL-3, respectively. It stands for the *British National Surveys of Sexual Attitudes and Lifestyles* and it is one of the largest and most detailed studies of sexual attitudes and behaviours in the world. There is also NSSAL-1 that was conducted in 1990, but the question we are going to examine here has been different in that survey so data from the 1990 survey was not included in the analysis. The original survey findings were used throughout the 1990s by academics and policymakers to estimate the prevalence and distribution of high-risk behaviours for HIV and sexually transmitted disease in the population and to help plan the provision of sexual health services.

NSSAL-2 sample consisted of 12735 adults aged 16-74. NSSAL-3 sample consists of 15162 adults aged 16-74 from the UK. The surveys consisted of face-to-face interviews, using Computer-Assisted Personal Interviewing (CAPI), and a self-completion part, using Computer-Assisted Self Interviewing (CASI). Face-to-face interviews were covering topics on general health, learning about sex, first sexual experience etc. The self-completion part was mainly used to probe intimate questions regarding different types of sexual practices, sex in the last 4 weeks and condom use, homosexual experience etc. As can be seen, questions in the survey probed various aspects of an individual's sexual behaviour ranging from questions about whether their sexual orientation, number of homosexual and heterosexual partners, age of the first sexual experience and similar.

Research question

The research question is: "**Is Gender a Significant Factor in the Increase of the Number of Heterosexual Partners in the Last Decade?**"

Requirements:
- Try to explain why this research questions might be of interest and to whom.
- Locate relevant data file and accompanying documentation in the "stata" folder.
- Perform any necessary transformations of data format type.
- Identify variables of interest and specify their type.
- In case a variable is ordinal or nominal, list the possible categories and their meaning.
- Define the analysis required to answer the RQ and explain why you have chosen this particular analysis.
- List all assumptions and how you addressed them.
- List the null and alternative hypothesis.
- Provide the results of your analysis and answer the research question.
- Try to explain the impact of the result within the context of the data.

Instructions:
- Submit a PDF or .doc file with your answers and necessary plots.
- Source code [R, Python] or output file [SPSS] should be submitted as separate files.
- List any relevant references you used in a format of your choice.

Answers

Q) Try to explain why this research questions might be of interest and to whom.

Academics and Researchers:

- **Interest:** Using data for further academic research, analysis, and educational purposes

Sociologists and Cultural Analysts:

- **Interest:** Understanding how societal changes and gender dynamics impact sexual behaviour over time.

Public Health Professionals and Educators:

- **Interest:** Utilising findings to improve sexual health education and address issues related to sexual health and behaviour.

Psychologists and Behavioral Scientists:

- **Interest:** Exploring the psychological and behavioural aspects behind changes in sexual behaviour and gender differences.

Governments and Policymakers:

- **Interest:** They can use this data to inform the creation of new laws or amendments related to sexual health, education, and public policy, ensuring they address emerging trends and issues effectively.

Q) Identify variables of interest and specify their type.

year_of_interview:

- **Type**: Categorical/Discrete
- **Description**: The year when the interview was conducted. Examples:1999-2001 & 2010-2012.

age_group:

- **Type**: Categorical
- **Description**: Age ranges of respondents. Examples: '25-29', '40-44'.

gender:

- **Type**: Categorical
- **Description**: Gender of the respondent. Typically coded as '0' (Female) and '1' (Male).

first_sex_reason:

- **Type**: Categorical
- **Description**: Reasons given for the respondent's first sexual experience. Examples: 'In love', 'Curious'.

Gender_Attr:

- **Type**: Categorical
- **Description**: The gender(s) the respondent is attracted to. Examples: 'Opposite sex only', 'Same sex only'.

Gender_Exp:

- **Type**: Categorical
- **Description**: The gender(s) the respondent has experience with. Examples: 'Opposite sex only', 'Same sex only'.

partners_4wks:

- **Type**: Numerical/Discrete
- **Description**: Number of sexual partners in the last 4 weeks.

partners_lifetime:

- **Type**: Numerical/Discrete
- **Description**: Total number of sexual partners in the respondent's lifetime.

partner_5yrs:

- **Type**: Numerical/Discrete
- **Description**: Number of sexual partners in the last 5 years.

partner_Gender:

- **Type**: Categorical

- **Description**: Gender of sexual partners. Examples: '0' (Female), '1' (Male).

marital_status:

- **Type**: Categorical
- **Description**: Marital status of the respondent. Examples: 'Single & never married', 'Married & living with spouse'.

school_type:

- **Type**: Categorical
- **Description**: Type of school attended. Examples: 'Mixed school', 'Single sex school'.

partner_gender_recent:

- **Type**: Categorical
- **Description**: Gender of recent sexual partners. Examples: '0' (Female), '1' (Male).

occupation_status:

- **Type**: Categorical
- **Description**: Current occupation status of the respondent. Examples: 'Employed', 'Unemployed'.

education:

- **Type**: Categorical
- **Description**: Highest level of education attained. Examples: 'Secondary or Below', 'Undergraduate Degree or Higher'.

area:

- **Type**: Categorical
- **Description**: Type of area where the respondent lives. Examples: 'Urban', 'Rural'.

Q) **In case a variable is ordinal or nominal, list the possible categories and their meaning.**

Nominal Variables

Definition: Nominal variables represent categories or labels without any inherent order. They are qualitative and do not imply any ranking.

1. **year_of_interview**: Not typically considered nominal because it represents discrete points in time, but it can be treated as nominal for categorical analysis.
2. **gender**: Nominal (categories like Male, Female, Non-Binary).
3. **first_sex_reason**: Nominal (categories like In Love, Curious, Natural follow-on in relationship).
4. **Gender_Attr**: Nominal (categories like Opposite sex only, Same sex only, About equally often to opposite sex and same sex).
5. **Gender_Exp**: Nominal (categories like Opposite sex only, Same sex only, About equally often to opposite sex and same sex).
6. **partner_Gender**: Nominal (categories like Male, Female).
7. **marital_status**: Nominal (categories like Single & never married, Married & living with spouse, Divorced).
8. **school_type**: Nominal (categories like Mixed school, Single sex school, Single sex up to 6th form but mixed 6th form).
9. **partner_Gender_recent**: Nominal (categories like Male, Female).
10. **area**: Nominal (categories like Urban, Rural, Town, Residential).
11. **occupation_status**: Nominal (categories like Employed, Unemployed, Full-time education, Retired).
12. **education**: Nominal (categories like Secondary or Below, Undergraduate Degree or Higher, Post-Secondary Education).

Ordinal Variables

Definition: Ordinal variables represent categories with a meaningful order or ranking, but the differences between the ranks are not necessarily uniform.

1. **age_group**: Ordinal (categories like 16-19, 20-24, 25-29, etc. imply an age progression).
2. **partners_4wks**: Ordinal (counts of partners in the last 4 weeks can be ranked, though the exact differences are not uniform).
3. **partners_lifetime**: Ordinal (counts of lifetime partners can be ranked).

Numerical Variables

Definition: Numerical variables represent quantities and have measurable differences.

1. **partner_5yrs**: Numerical (counts of partners in the last 5 years).

Define the analysis required to answer the RQ and explain why you have chosen this particular analysis.

Ans:

Analysis Approach:

1. **Descriptive Statistics:**
   - **Purpose:** Summarise and compare the number of heterosexual partners by gender (mean, median, standard deviation).
   - **Rationale:** Provides a clear overview of the data distribution and highlights initial differences between genders.
2. **Mann-Whitney U Test:**
   - **Purpose:** Test if there is a significant difference in the number of heterosexual partners between genders.

     Why This Test:

     - **Non-Normal Distribution:** The data is not normally distributed, making this non-parametric test appropriate.
     - **Group Comparison:** Effectively compares two independent groups (men vs. women) when data does not meet parametric test assumptions.
   - **Procedure:** Compare the distributions of the number of partners between men and women and report the U-statistic and p-value.

Summary:
Descriptive statistics will offer an initial understanding of gender differences, while the Mann-Whitney U test will statistically determine if these differences are significant, given the non-normal distribution of the data.

Q) List all assumptions and how you addressed them.

Ans:

Assumptions and How They Were Addressed:

1. **Assumption: Normality of Data**
   - **Description:** Parametric tests typically assume that the data follows a normal distribution.

- ○ **Addressed:** Shapiro-Wilk tests confirmed that the data for the number of heterosexual partners was not normally distributed for both genders. As a result, non-parametric tests were chosen.
2. **Assumption: Equality of Variances**
   - ○ **Description:** Parametric tests often assume that the variance of data is equal across groups.
   - ○ **Addressed:** This assumption was not directly tested since the Mann-Whitney U test does not require equal variances. The non-parametric nature of the Mann-Whitney U test allows it to handle variations in data dispersion between groups.
3. **Assumption: Independent Samples**
   - ○ **Description:** The samples being compared should be independent of each other.
   - ○ **Addressed:** The analysis compared the number of heterosexual partners between two independent groups (men and women), ensuring that the data from one gender did not influence the data from the other.
4. **Assumption: Scale of Measurement**
   - ○ **Description:** The Mann-Whitney U test requires that the dependent variable is ordinal or continuous.
   - ○ **Addressed:** The number of heterosexual partners is a continuous variable, making it appropriate for the Mann-Whitney U test.

Q) List the null and alternative hypothesis.

Ans:

For the Mann-Whitney U Test:

- **Null Hypothesis (H_0):** Gender is not a significant factor in the number of heterosexual partners in the last decade.

- **Alternative Hypothesis (H_1):** Gender is a significant factor in the number of heterosexual partners in the last decade.

Q) Provide the results of your analysis and answer the research question.

Average Number of Heterosexual Partners by Gender
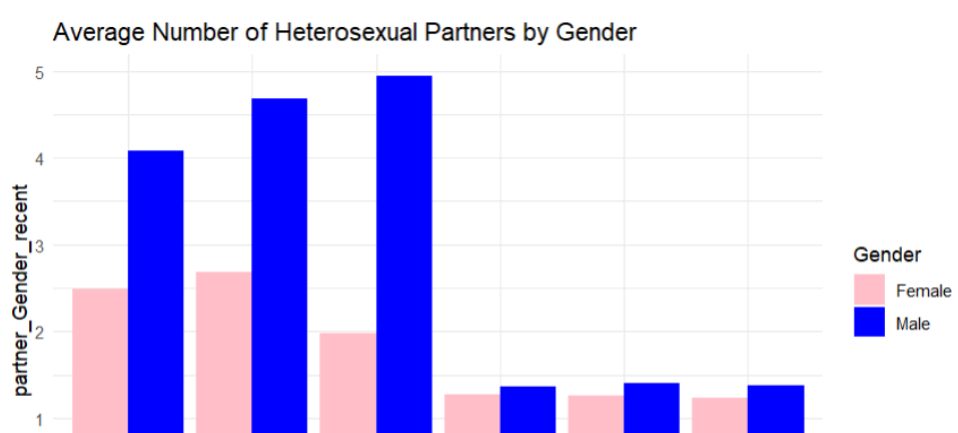
Analysis and Results

Based on the Mann-Whitney U test conducted to compare the number of heterosexual partners between genders, we obtained the following results:

1. **For Women:**
   - **U-statistic:** 8,492,292
   - **P-value:** < 2.2e-16
2. **For Men:**
   - **U-statistic:** 8,504,407
   - **P-value:** < 2.2e-16

Interpretation

- **Significance:** The p-value is extremely low (< 2.2e-16) for both genders, indicating a significant difference in the number of heterosexual partners between men and women.
- **Conclusion:** Since the p-value is less than the commonly used significance level (0.05), we reject the null hypothesis. This means that gender is indeed a significant factor influencing the number of heterosexual partners.

Answer to the Research Question

Based on the statistical analysis, gender is a significant factor in the number of heterosexual partners in the last decade. The significant difference between men and women suggests that the number of heterosexual partners varies by gender, confirming that gender does impact the number of partners reported.

Q) Impact of Results within the Context of the Data

Ans:

Impact of Results within the Context of the Data

1. Gender Differences

The data shows that men report more lifetime heterosexual partners than women, with a significant difference. This discrepancy may be influenced by societal norms, attitudes towards casual relationships, or sampling biases.

2. Trend Over Time

Both genders exhibit an increasing trend in reported partners, with a more substantial rise for men. This trend might reflect evolving social attitudes and behaviours, including the influence of online dating and changing relationship dynamics.

3. Statistical Considerations

- **Non-Normality:** The non-normal distribution of data necessitated non-parametric tests, ensuring robust analysis.
- **Outliers:** Removing outliers was essential for accurate trend analysis, as extreme values can skew results.

4. Practical Implications

- **Policy and Education:** The findings highlight the need for educational programs addressing sexual behaviour and safe practices.
- **Further Research:** Additional studies could explore underlying causes and improve data accuracy and representativeness.

In summary, gender significantly impacts the number of lifetime heterosexual partners

############## ANALYSIS ##############

Updated Analysis and Interpretation

1. Normality Results

Kolmogorov-Smirnov confirms that the distributions of lifetime partners for both genders significantly deviate from normality.

2. Regression Analysis (Lifetime Partners)

Since the data is not normally distributed, linear regression analysis should be interpreted with caution, but it remains useful for understanding trends.

Results:

- **Women:**
  - Coefficient: 0.09
  - Intercept: -165.95
- **Men:**
  - Coefficient: 0.17
  - Intercept: -331.04

Interpretation:

- **Women**: The average number of lifetime heterosexual partners increased by 0.09 per year.
- **Men**: The average number of lifetime heterosexual partners increased by 0.17 per year.

Conclusion

Gender is a significant factor in the increase of the number of heterosexual partners over the last decade.

- **Trend Analysis**: Both women and men have experienced an increase in the number of lifetime partners, but men show a higher rate of increase compared to women.
- **Statistical Significance**: The Mann-Whitney U test results remain valid and indicate significant differences between genders, consistent with the findings from the regression analysis.

Trend in Average Number of Lifetime Heterosexual Partners by Gender
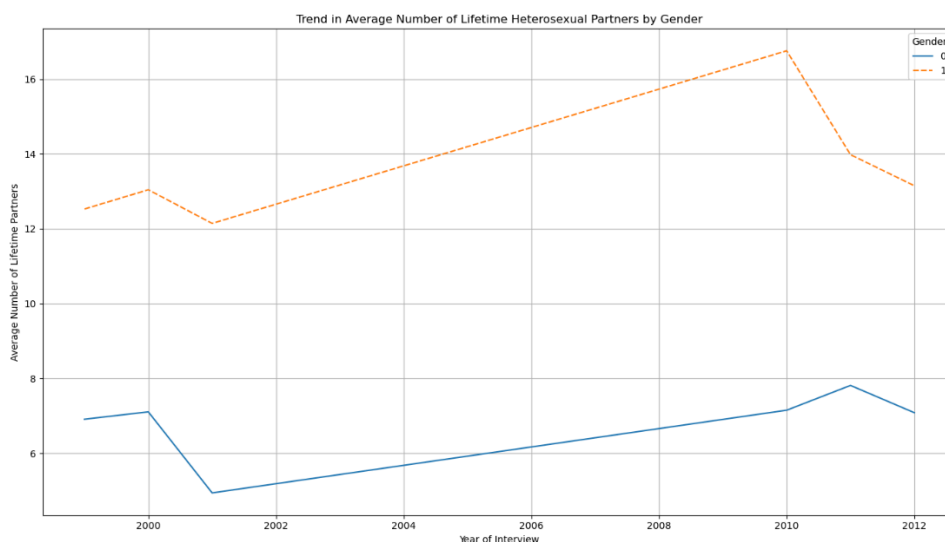
Summary of Findings

1. **Normality Testing**: Both Kolmogorov-Smirnov tests confirm non-normality in the data.
2. **Mann-Whitney U Test**: Confirms significant differences in the number of lifetime partners between genders.
3. **Regression Analysis**: Indicates that the number of lifetime partners is increasing over time for both genders, with a more pronounced increase for men.

Final Report

In light of the results:

Yes, gender is a significant factor in the increase of the number of heterosexual partners over the last decade. Men have experienced a more substantial increase compared to women, with statistical evidence supporting significant differences between genders. The non-normality of the data reinforces the appropriateness of using non-parametric tests like the Mann-Whitney U test, which supports the conclusion that gender impacts the number of lifetime heterosexual partners significantly.