

Data Review Assignment

Data Set Profile

Data Set Profile

2a. Who is credited as the creator and/or contributors of this data set?

Our Professor is the creator of the data set pertaining to this assignment. The tool she used to gather said data is called "Amazon mechanical turk". Each person ended up getting paid a grand total of fifty cents. However, the creator of the survey, our professor, did not know which individuals

The contributors are the people providing the information that is in the data set.

2b. Who are they?

These are people who identify as gay or lesbian and are above the age of eighteen and below the age of eighty-nine

2c. What are the sources of their data? Why did they create or comp

She wanted to better understand how gay and lesbian friends ended up gaining their identity. She asked each individual to use five terms to describe their specific group they associated with within the survey and then piped the data to formulate a series of questions to ask the individuals with quantitative outputs on a 1-5 scale about how they felt about certain general assumptions about their group. It was also one of the final research studies our profession had to do for her dissertation prior to her receiving her doctorate.

This Data has been presented at national conferences and has been used in many professions. Also it has been submitted and is currently under review

2d. How has it been used

Data Set Evaluation

Analyze the structure and potential biases in the data:

3a. How is the data structured?

It is based on a survey data structure requiring active input from the participants. Each row represents a single participant with a total of 1,285 cases, and each column corresponds to a specific variable. The datasets include demographic information such as age, ethnicity, and education as well as psychological labels and their corresponding values. Some variables can be categorized such as Ethnicity or education and others are measured on continuous or ordinal scales such as self esteem and private regard.

3b. What fields have been chosen?

Age, ethnicity, education, self esteem, private regard, sexual identification, satisfaction with life, and sex role inventory.

3c. How might this impact its use?

By including both demographic and psychological variables, the dataset allows for analyses that examine how demographic factors such as ethnicity and sexual identification can correspond to different psychological outputs.

3d. Have the creators described their data cleaning process?

By taking out things they didn't need such as unnecessary identifying information i.e. IP addresses. They also discarded something known as text data and manually took all of similar entities and standardized them. She did it by alphabetizing them and then putting them into categories such as "capitalization", "different spelling", etc. this is just to sift for answers to very specific questions. Lastly she was checking to identify any reverse-coded entities which are basically any text

3e. What effect might their choices have on the data?

It can make the data more accurate, but also it can take away some key informative details at an alternate expense.

3f. How might the creators' identities and goals have shaped the data, either intentionally or inadvertently?

Because she is an ally of the people surveyed for this date. However, an inadvertent affect this creator could have potentially posed to this data may be that as a CIS gendered person she may not understand the full scope of what some demographics face or the terminology due to their not having first-hand experience in said communities.

3g. What would you use this data for?

Potentially I could use it for looking at how specific demographics under the LGBTQ+ feel pertaining to certain social stigmas in certain regions.

Statistical Analysis (Using SPSS)

Conduct the following analyses on the data set:

4a/4b. **Descriptive Statistics:** Report means, standard deviations, minimums, maximums, skewness, and kurtosis for key demographic and psychological variables.

Psychological variables:

Self-Esteem

- **Mean:** 3.03
- **Standard Deviation:** 0.63

- **Minimum:** 1.00
- **Maximum:** 4.00
- **Median:** 2.5

Private Regard (personal feelings)

- **Mean:** 5.81
- **Standard Deviation:** 1.05
- **Minimum:** 1.71
- **Maximum:** 7.00
- **Median:**

Satisfaction with Life

- **Mean:** 4.73
- **Standard Deviation:** 1.53
- **Minimum:** 1.00
- **Maximum:** 7.00
- **Median:** 5

Demographic variables:

Age

- **Mean:** Not explicitly stated, but frequency distribution is provided.
- **Standard Deviation:** Not explicitly stated.
- **Minimum Age:** 18
- **Maximum Age:** 63

Ethnicity

Caucasian (15%)

African-American (1.4%)

Hispanic (1.2%)

Education

- **Graduate School:** 2.8%
- **4 years of college:** 7.1%
- **3 years of college:** 2.3%
- **2 years of college:** 3.3%
- **1 year of college:** 1.6%
- **High school diploma/GED:** 3.0%

Sex Role Inventory

- **Mean:** 4.15
- **Standard Deviation:** 0.54
- **Minimum:** 2.17
- **Maximum:** 5.00
- **Median:** 3.585

Sexual Identification

- **Mean:** 3.95
- **Standard Deviation:** 0.85
- **Minimum:** 1.00
- **Maximum:** 5.00
- **Median:** 3

Start: 2/17/25

Due: 2/23/25

Note: I put "sex role inventory" and "sexual identification under demographics" as it is a societal classification