Gender and Mental Depression: Analyzing Census Data Through Statistical Methods*

Christina Nguyen   Marcus Barnes   Nibras Ar Rakib

October 21, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

We used the Medical Expenditure Panel Survey dataset from the Integrated Public Use Microdata Series, which contains microdata from different time points related to U.S. healthcare expenditures (Blewett et al. 2023). We performed all analysis using a popular statistical language called R (R Core Team 2023). Our dataset contains age, sex, marital status, race, highest level of education attainment, total personal income, and K6 score (Kessler et al. 2012).All the variable definitions can be found in the Appendix A.

Our goal was to identify people, based on gender and education, who are suffering from non-specific psychological distress based on the Kessler scale (Kessler et al. 2012) before and after the COVID-19 pandemic.

Our goal was to identify respondents, based on gender and education, who were suffering from non-specific psychological distress based on the Kessler scale before and after the COVID-19 pandemic. To support our analyses, we split into two datasets for 2018 and 2021. We also excluded the respondents whose ages are below 19 and over 80. After enforcing our inclusion criteria, we had 15716 respondents for 2018 and 12685 for 2021. To give a preview of our data, Table 1 and Table 2 show the first five rows for each row for 2018 and 2021, respectively.

Table 1: Year 2018

year	age	sex	marstat	racea	educ	inctot	k6sum
2018	27	2	10	100	400	32000	3
2018	34	2	10	100	201	25000	4
2018	39	1	10	100	201	30000	0
2018	36	2	10	100	501	30217	0
2018	30	1	10	100	301	31644	0

Table 2: Year 2021

year	age	sex	marstat	racea	educ	inctot	k6sum
2021	59	2	10	100	201	9508	2
2021	73	1	10	100	201	25508	2
2021	21	1	50	100	115	20000	2
2021	44	2	10	100	504	1500	1
2021	41	1	10	100	201	41500	1

2.2 Measurement

2.2.1 Gender and Mental Health

Research consistently shows that women report higher rates of depression than men (CITE). Biological, social, and psychological factors contribute to this disparity. Women may experience unique stressors, such as role strain from balancing work and family obligations, as well as being subject to gender-based discrimination (CITE). These pressures can exacerbate mental health issues, leading to higher rates of depression among women, particularly those in lower socioeconomic or educational strata.

2.2.2 Educational Attainment and Mental Health

Education is often seen as a protective factor against mental health issues. Studies suggest that higher levels of education correlate with improved mental health outcomes (CITE). Educated individuals tend to have better access to mental health resources, higher socioeconomic status, and greater job security, all of which contribute to lower rates of depression (CITE). Conversely, those with lower education levels may face more financial strain, reduced employment opportunities, and limited access to healthcare, increasing their vulnerability to depression.

2.2.3 Intersection of Gender, Education, and Depression

Despite the protective nature of education, the gender gap in mental health persists even at higher levels of educational attainment. Some studies suggest that women with high education levels still experience depression due to external stressors, including workplace discrimination and societal expectations (CITE). Additionally, the mental health benefits of education may differ between genders, with men benefiting more from the social status and economic opportunities that education provides, while women may continue to face challenges related to gender roles (cite).

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table **??**.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

A.1 Variable Definitions

Variable: "AGE"

Name:	AGE
Label:	Age
Variable Text:	AGE reports the individual's exact age, calculated from date of birth, as of the last day (12/31) of the survey year. Date of birth and age were asked for each reporting unit member, and then exact age was calculated from date of birth. Where the calculated age and the age provided did not match, inconsistencies were reviewed and resolved. When date of birth was not provided, but age was provided, the month and year of birth were assigned randomly from among the possible valid options. For any cases still not accounted, age was imputed using: (1) the mean age difference between MEPS participants with certain family relationships (where available) or (2) the mean age value for MEPS participants.
Concept:	Core Demographic Variables -- PERSON
Start Position:	99
End Position:	101
Width:	3
Variable Format:	numeric
Implied Decimal Places:	0

Coder Instructions:	CodesAGE is a 3-digit-numeric variable.
	085: Top code for 85 years or older (2001-forward)
	090: Top code for 90 years or older (1996-2000)
	996: Missing

Variable: "Sex"

Name:	SEX
Label:	Sex

Variable Text:	SEX indicates whether the person was male or female.
	Collection of information on SEX in MEPS Data on the sex of each reporting unit (RU) member was determined during the NHIS interview, and was then verified, and, if necessary, corrected during each MEPS interview.
	If the respondent was a new RU member or their sex was not ascertained in the NHIS interview, MEPS initially used the first name of the respondent to assign their sex. If the first name gave no clear indication of sex, the reported family relationships were used to assign sex. If the sex of the respondent was still unclear, sex was randomly assigned.
	The NHIS method of ascertaining the sex of the respondent, which primarily informs the MEPS version of SEX, is similar to the MEPS method. First, sex of the respondent was inferred from the individual's first name or family relationships. If the sex of the respondent was unclear, the interviewer was instructed to explicitly ask the person's sex. Beginning in 1998, interviewers were told to "enter your best guess" when the respondent either did not know or refused to answer the direct question that was asked about the person's sex.
Concept:	Core Demographic Variables -- PERSON
Start Position:	102
End Position:	102
Width:	1
Variable Format:	numeric
Implied Decimal Places:	0
Categories	

Value	Label
1	Male
2	Female
7	Unknown-refused
8	Unknown-not ascertained
9	Unknown-don't know

Variable: "MARSTAT"

Name:	MARSTAT
Label:	Legal marital status
Variable Text:	For persons age 16 and older, MARSTAT reports the person's legal marital status.
Concept:	Core Demographic Variables -- PERSON
Start Position:	103
End Position:	104
Width:	2
Variable Format:	numeric
Implied Decimal Places:	0
Categories	

Value	Label
00	NIU
10	Married
20	Widowed
30	Divorced
40	Separated
50	Never married
99	Unknown marital status

Variable: "RACEA"

Name:	RACEA
Label:	Main Racial Background (Pre-1997 Revised OMB Standards), self-reported or interv
Variable Text:	RACEA reports the race of the respondent. If not ascertained, the race and/or ethn
Concept:	Ethnicity/Nativity Variables -- PERSON
Start Position:	105
End Position:	107
Width:	3

Variable Format: numeric
Implied Decimal Places: 0
Categories

Value	Label
100	White
200	Black/African-American
300	Aleut, Alaskan Native, or American Indian
310	Alaskan Native or American Indian
320	Alaskan Native/Eskimo
330	Aleut
340	American Indian
350	American Indian or Alaskan Native and any other group
400	Asian or Pacific Islander
410	Asian
411	Chinese
412	Filipino
413	Korean
414	Vietnamese
415	Japanese
416	Asian Indian
420	Pacific Islander
421	Hawaiian
422	Samoan
423	Guamanian
430	Other Asian or Pacific Islander
431	Other Asian or Pacific Islander (1992-1995)
432	Other Asian or Pacific Islander (1996)
433	Other Asian or Pacific Islander (1997-1998)
434	Other Asian (1999 forward)
500	Other Race
510	Other Race (1963-1977)
520	Other Race (1978)
530	Other Race (1979-1991)
540	Other Race (1992-1995)
550	Other Race (1996)
560	Other Race (1997-1998)
570	Other Race (1999-2002)
580	Primary Race not releasable
600	Multiple Race, No Primary Race Selected

610	Multiple Race, including Asian, excluding Black and White
611	Multiple Race, including Asian and Black, excluding White
612	Multiple Race, including Asian and White, excluding Black
613	Multiple Race, including Black, excluding Asian and White
614	Multiple Race, including Black and White, excluding Asian
615	Multiple Race, including White, excluding Asian and Black
616	Multiple Race, including Asian, White, and Black
617	Multiple Race, excluding Asian, White, and Black
900	Unknown
970	Unknown-refused
980	Unknown-not ascertained
990	Unknown (1997 forward: Don't know)

Variable: "MARSTAT"

Name:	MARSTAT
Label:	Legal marital status
Variable Text:	For persons age 16 and older, MARSTAT reports the person's legal marital status.
Concept:	Core Demographic Variables -- PERSON
Start Position:	103
End Position:	104
Width:	2
Variable Format:	numeric
Implied Decimal Places:	0
Categories	

Value	Label
00	NIU
10	Married
20	Widowed
30	Divorced
40	Separated
50	Never married
99	Unknown marital status

Variable: "EDUC"

Name:	EDUC

Label:	Educational attainment
Variable Text:	EDUC reports the highest level of schooling an individual had completed, in terms of completed grades for persons with less than a high school degree, and in terms of degrees attained for high school graduates and those with higher education. EDUC is available for all survey participants age 5 and older at the time of their first MEPS interview. It is an IPUMS MEPS constructed variable.
	See Comparability section for more information on the construction of EDUC.
	Education Variables offered through IPUMS MEPS In addition to EDUC, there are several other educational attainment variables offered through MEPS. Unlike EDUC, these variables are not available for every MEPS year.
	EDUCYR: Years of education completed (available 1996-2011 and 2014 forward)
	HIDEG: Highest degree completed (available 1996-2011 and 2014 forward)
	EDUYRDG: Years of education and highest degree completed (available 2011-2015)
	EDRECODE: Years of education, recode (available 2011-2015)
Concept:	Education Variables -- PERSON
Start Position:	108
End Position:	110
Width:	3
Variable Format:	numeric
Implied Decimal Places:	0
Categories	

Value	Label
000	NIU
100	Grade 12 or less, no high school diploma or equivalent
101	Grade 8 or less (no further detail)
102	Never attended/kindergarten only
103	Grades 1-11 (no further detail)
104	Grade 1
105	Grade 2
106	Grade 3
107	Grade 4
108	Grade 5
109	Grade 6
110	Grade 7
111	Grade 8
112	Grade 9-12, no diploma (no further detail)
113	Grade 9
114	Grade 10
115	Grade 11
116	12th grade, no diploma
200	High school diploma or GED
201	High school graduate
202	GED or equivalent
300	Some college, no 4yr degree
301	Some college, no degree
302	AA degree: technical/vocational/occupational
303	AA degree: academic program
400	Bachelor's degree (BA,AB,BS,BBA)
500	Master's, Professional, or Doctoral Degree
501	Master's degree (MA,MS,Med,MBA)
502	Professional (MD,DDS,DVM,JD)
503	Doctoral degree (PhD, EdD)
504	Other degree
505	Professional School or Doctoral degree, topcoded (MD, DDS, DVM, JD, PhD, EdD)
996	No degree, years of education unknown
997	Unknown--refused
998	Unknown--not ascertained
999	Unknown--don't know

Variable: "INCTOT"

Name:	INCTOT

Label: Total personal income

Variable Text:	INCTOT reports the sum of all person-level income for the current calendar year, excluding income from tax refunds and capital gains.

INCTOT includes annual earnings from wages, salaries, bonuses, tips, commissions; business and farm games and losses; unemployment and workers' compensation; interest and dividends; alimony, child support, and other private cash transfers; private pensions, IRA withdrawals, social security, and veterans payments; supplemental security income and cash welfare payments from public assistance, Temporary Assistance for Needy Families, and related programs; gains or losses from estates, trusts, partnerships, S corporations, rent, and royalties; and a small amount of "other" income. Person-level income excluded tax refunds and capital gains.

Logical editing or weighted, sequential hot-deck imputation was used to impute income amounts for missing values (both for item non-response and for person in the full-year file who were not in round 3). Reported income components were generally left unedited.

Related Variables
INCWAGE: annual wage and salary income of individuals

INCBUS: annual business income of individuals

INCUNEMP: annual unemployment compensation income of individuals

INCWKCOM: annual workers' compensation income of individuals

INCINT: annual interest income of individuals

INCDIVID: annual dividend income of individuals

INCRETIR: annual pension income of individuals

Concept:	Total Income and Earnings Variables -- PERSON
Start Position:	111
End Position:	119
Width:	9
Variable Format:	numeric
Implied Decimal Places:	2
Coder Instructions:	CodesINCTOT is a 9-digit numeric variable with 2 implied decimals. That is, values of 012345678 should be interpreted as 123456.78. The command files delivered with IPUMS extracts automatically divide INCTOT by 100, so no further adjustment is needed.

9999999.96: Not in Universe
9999999.97: Unknown-refused
9999999.98: Unknown-not ascertained
9999999.99: Unknown-don't know |

Variable: "K6SUM"

Name:	K6SUM
Label:	K6 score for nonspecific psychological distress: last 30 days

Variable Text:	For persons eligible for the self-administered questionnaire (SAQELIG), K6SUM is the summed scale value measuring nonspecific psychological distress over a 30-day recall period. This scale, developed by Ronald C. Kessler and known as the Kessler 6 Scale (K6), asks about six manifestations of nonspecific psychological distress.

Kessler recommends scoring the scale by assigning 0 to 4 points for each of the six questions, based on the reported frequency of the feelings (i.e., 0 for "none of the time"; 1 for "a little of the time"; 2 for "some of the time"; 3 for "most of the time"; and 4 for "all of the time"). The range for summed responses on the K6 Scale is thus 0 to 24, with 0 suggesting the lowest level of nonspecific psychological distress, and 24 suggesting the highest level of nonspecific psychological distress. According to the scoring criteria proposed by Kessler, persons with a score of 13 or greater are likely to be experiencing severe mental illness.

Kessler's instrument asks how often, during the past 30 days, the respondent felt:

So sad that nothing could cheer you up? (ASAD)

Nervous? (ANERVOUS)

Restless or fidgety? (ARESTLESS)

Hopeless? (AHOPELESS)

That everything was an effort? (AEFFORT)

Worthless? (AWORTHLESS)
As noted above, acceptable responses fell into five categories, ranging from "none of the time" to "all of the time."

Concept:	Adult Mental Health Variables -- PERSON
Start Position:	120
End Position:	121
Width:	2
Variable Format:	numeric
Implied Decimal Places:	0
Categories	

Value	Label
96	NIU
98	Unknown-not ascertained

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

Examining how the model fits, and is affected by, the data

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

Checking the convergence of the MCMC algorithm

References

Blewett, Lynn A., Julia A. Rivera Drew, Daniel Backman, Annie Chen, Grace Cooper, Megan Schouweiler, Stephanie Richards, and Michael Westberry. 2023. "IPUMS Health Surveys: Medical Expenditure Panel Survey: Version 2.3." Minneapolis, MN: IPUMS. https://doi.org/10.18128/D071.V2.3.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data.* https://doi.org/10.5281/zenodo.3960218.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. T. Normand, E. E. Walters, and A. M. Zaslavsky. 2012. "Kessler Psychological Distress Scale." https://doi.org/10.1037/t08324-000.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.