**Project 1: Demographic testing**

**Submission Rules**

1. This is an **individual project**. While you are welcome to ask for help from the instructor(s) and teaching assistants, you are expected to complete the data analysis (including coding) and write-up of the report on your own.
2. Submit your project as a MS Word or PDF file to Canvas by 5:00 pm Central US time on Sunday, November 19, 2023. All other forms will not receive credit.
3. Submit your project code separately as a .sas, .r, or .RMD file in the Project 1 code assignment. The file name for this code must include your name or initials.
4. Reports turned in up to 24 hours late without first requesting an extension by Saturday will be deducted 20% of total possible project grade (10.6 points). Reports more than 24 hours late will not be accepted.
5. This project assignment should **not** be part of your submission.

**The report:**

1. In your report, share **only** statistical summaries or inferential results or graphs that further your argument, making sure to give **titles to all tables and graphs/figures (example: Table 1 or Figure 1).**

* Use tables (including the ones presented below) when you need to summarize many different statistics. All tables should be created in your word document.  Provide only the information needed for your report.  Make sure to round statistical summaries to two decimal places or three significant figures, and either write out the number or use scientific notation (example: 0.00032 or 3.2 x 10-4).  Tables that resemble those from software will lose points.
* Make sure you label graph axes, including putting the unique title information in the title of the graph (and not as a note below the graph). You are allowed to use graphs directly from software, but please edit them to make sure a person can understand what is being visualized in the graph.

1. Write your report with the goal of it being **no longer than 3 pages** (excluding any figures or tables).  Focus on brevity over the page limit, and try for 1500 words or less.  A report that exceeds 3 pages but is 12 point font with clear headings on the sections and paragraphs will not lose points.  A report that attempts to be less than 3 pages in 9 point font with no titles may lose points for clarity.
2. Write your report as you would for a client or collaborator, in full sentences and paragraphs. Make sure your presentation of your work is clean, readable, and professional. **Two points** will be deducted for excessively sloppy submissions.

**The code:**

1. You should provide your SAS or R code as a .sas or .r file in the separate assignment and give it the title “Your Name Project 1 code”. This code will be checked only if a statistic, graph, or inference does not match what was expected. You do not have to annotate your code, but make sure every line of code is exactly as you used it for the analysis. This code is worth **one point.**

**Background**

Every four years, Hennepin County Public Health conducts a survey about residents’ mental and physical health, as well as the social determinants of health. To learn more, please see their website: <https://www.hennepin.us/your-government/research-data/shape-surveys>

We are going to work with a cleaned subset of the 2022 survey, focusing on those individuals who have had one or more days where their mental health was not good in the 31 days before the start of the survey.

For your final project, you will test if two individual characteristics are associated with or are predictors of a **resident’s number of not good mental health days** (MHdays). The two predictors are:

1. **Insecurity score** (Insecurity2): a sum score of 4 questions about food insecurity, 2 questions about housing insecurity, 2 questions about safety in their community, and 1 question about transportation insecurity. A higher score means more insecurity in their lives.
2. **Discrimination in health care** (Discrim2): Whether or not the resident felt they were treated unfairly or discriminated against when seeing medical, mental, or dental care.

NOTE: both the number of not good mental health days and the insecurity score are count data, but for this work, treat them as if they were continuous numerical variables.

**Project 1: Demographics**

Medical papers often start with a giant table of demographics (which is called Table 1.) Testing demographics can give insight into possible **confounders**, or reasons why our explanatory and response variables appear related. If a variable is significantly associated with **both** a predictor and the outcome, it may be a confounder.

For this report, we will be examining the relationship between three demographic variables (Gender, Trans status, and Race/ethnicity) and our potential predictors, and the relationship between those demographic variables and our outcome of interest. You will be performing 9 tests (and creating confidence intervals, where appropriate): 3 for each predictor (Insecurity score and Discrimination in health care), and 3 for the outcome (Number of not good mental health days).

**Introduction: (2 points)**

* Give an overview of the research for this analysis
* Give detail about what you will be testing and why.

**Testing for associations between demographics and the Insecurity predictor. (15 points)**

You will be testing the demographic variables Gender3, TransGender, and EthRace5\_2022 versus the Insecurity score (Insecurity2).

For each demographic:

* Find the number of people surveyed who are in each demographic category. **Enter those values into table 1.**
* Find the average Insecurity score for each category of the demographic. **Enter those values into table 1.**
* Find the standard deviation of the Insecurity score for each category of the demographic. **Enter those values into table 1.**
* Using the summary statistics and any graphs you choose to make, determine what test to perform to see if there is a difference in the average Insecurity score for at least one category. **List what test you have chosen to perform in your text and explain why.**
* Perform the test chosen and **enter the p-value for your test in table 1.**
* If the test results in a significant difference for one or more groups, **enter the groups that are significantly different (and which one is bigger) in table 1 (provide only those pairwise comparisons that are significantly different).**

**Table 1: Demographics versus the insecurity score**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable/Categories** | **n** | **Mean** | **SD** | **p-value** | **Differences1** |
| **Gender** |  |  |  |  |  |
| Male |  |  |  |  |  |
| Female |  |  |  |  |
| Agender/Genderqueer/Nonbinary |  |  |  |  |
| **Trans status** |  |  |  |  |  |
| Cis |  |  |  |  |  |
| Trans |  |  |  |  |
| **Race/Ethnicity** |  |  |  |  |  |
| Native 2022 inclusive |  |  |  |  |  |
| Hispanic |  |  |  |  |
| Asian-NH |  |  |  |  |
| Black-NH |  |  |  |  |
| White-NH |  |  |  |  |

1Include which groups are significantly different, if any, and if they are different, which group is bigger. If there are no significant differences, write “None”

**Testing for associations between demographics and the Discrimination in health care predictor (15 points)**

You will be testing the demographic variables Gender2, TransGender, and BIPOCAI versus the Discrim2 variable.

For each demographic:

* Find the relative risk of being treated unfairly or discriminated against in medical, mental, or dental care for those who have the characteristic vs those who do not. **Enter the relative risk in table 2.**
* Find the 95% confidence interval for that relative risk.
  + **Enter the lower value of the interval into table 2.**
  + **Enter the upper value of the interval into table 2.**
* Using the summary statistics and any graphs you choose to make, determine what test to perform to see if there is a difference in the risk for the two groups (i.e. whether the two variables are independent or associated.) **List what test you have chosen to perform in your text and explain why.**
* Perform the test chosen and **enter the p-value for your test in table 2** (note: your p-value should be from running the separate test, not from your relative risk calculations)**.**

**Table 2: Demographic risk of discrimination in health care**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable/Categories** | **Relative Risk2** | **95% confidence interval for RR** | **p-value** |
| Gender2 |  |  |  |
| Agender/Genderqueer/ Nonbinary |  |  |  |
| Male/Female | 1 | N/A |  |
| Trans Status |  |  |  |
| Trans |  |  |  |
| Cis | 1 | N/A |  |
| BIPOC/AI |  |  |  |
| BIPOC/American Indian |  |  |  |
| white | 1 | N/A |  |

2 The relative risk of being treated unfairly or discriminated against in medical, mental, or dental care vs not

**Testing for associations between demographics and the number of not good mental health days outcome. (15 points)**

You will be testing the demographic variables Gender3, TransGender, and EthRace5\_2022 versus the number of not good mental health days (MHdays).

For each demographic:

* Find the number of people surveyed who are in each demographic category. **Enter those values into table 3.**
* Find the average number of not good mental health days for each category of the demographic. **Enter those values into table 3.**
* Find the standard deviation of number of not good mental health days for each category of the demographic. **Enter those values into table 3.**
* Using the summary statistics and any graphs you choose to make, determine what test to perform to see if there is a difference in the average value for at least one category.  **List what test you have chosen to perform in your text and explain why.**
* Perform the test chosen and **enter the p-value for your test in table 3.**
* If the test results in a significant difference for one or more groups, **enter the groups that are significantly different (and which one is bigger) in table 3 (provide only those pairwise differences that are significantly different).**

**Table 3: Demographics versus number of not good mental health days**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable/Categories** | **n** | **Mean** | **SD** | **p-value** | **Differences1** |
| **Gender** |  |  |  |  |  |
| Male |  |  |  |  |  |
| Female |  |  |  |  |
| Agender/Genderqueer/Nonbinary |  |  |  |  |
| **Trans status** |  |  |  |  |  |
| Cis |  |  |  |  |  |
| Trans |  |  |  |  |
| **Race/Ethnicity** |  |  |  |  |  |
| Native 2022 inclusive |  |  |  |  |  |
| Hispanic |  |  |  |  |
| Asian-NH |  |  |  |  |
| Black-NH |  |  |  |  |
| White-NH |  |  |  |  |

1Include which groups are significantly different, if any, and if they are different, which group is bigger. If there are no significant differences, write “None”

**Conclusion: (3 points)**

Discuss the results.

* Did any of the demographic variables show an association with a predictor (Insecurity score or Discrimination in health care) or the outcome (Number of not good mental health days)?
* If so, did any variables have an association with **both** a predictor and the outcome?
* What does this mean for the potential of any of these demographics being a confounder?