

Stat 21 Test 1

Your name here

Due: March 4, 2020 by noon EST

This test is due on to be submitted on Gradescope on **March 4** by **12:00pm EST**. Please use the **#test1_questions** channel on Slack to post any clarification questions. Do not ask questions like “Is [this] the right answer?”

You must submit your solutions as a single **PDF** document uploaded to **Gradescope**. You may use R markdown to write up your solutions alone or you may use R markdown and hand-written solutions. **You must show all of your work**, including code input and output. Please make sure each problem is **clearly labeled** and that any handwritten components (such as pictures or equations) are easily readable in the PDF document. You may want to use a service like CamScanner (<https://www.camscanner.com/>) to help you upload handwritten pages and Small PDF (<https://smallpdf.com/merge-pdf>) to merge multiple PDFs into a single document.

You are permitted to reference all class material and use the internet (though I am not sure it will be very helpful). You are not permitted however, to get assistance from any other person, online or otherwise.

- Your file should contain the code to answer each question in its own code block. Your code should produce plots/output that will be automatically embedded in the output pdf file.
- Each answer must be supported by written statements and relevant plots.
- In order to knit this document, make sure you have installed the following packages in your version of RStudio: `ggplot2`, `tidyverse`, `gridExtra`, `knitr`

Gender discrimination

For problems 1-2¹, consider the following setting. A large, global supermarket chain is facing a class action lawsuit for gender discrimination in their job promotion practices. Each store location employs people at two levels: a general floor worker level and a supervisory level. The claim of the lawsuit is that whether or not an employee is promoted to a supervisory level depends on that employee's gender identity. To investigate this claim, a random sample of 100,006 supermarket employees is obtained from all stores in the US.

Problem 1

One of the prosecutors suggests collecting the data on the sample of 100,006 employees by categorizing whether or not the employees identified as women. Out of the 27,604 supervisors surveyed, 12,005 identified as women and out of the 72,402 floor workers surveyed, only 31,133 identified as women.

Determine whether or not there is evidence of discrimination. Use an $\alpha = 0.01$ level of significance. Show all your work and interpret your conclusion within the context of the problem and the appropriateness of the required assumptions.

Problem 2

Another prosecutor has recently taken a statistics class and has a different idea. They suggest that instead of categorizing gender according to a binary system, they take extra privacy precautions and invest some additional money into interviewing the random sample of 100,006 employees to instead ask participants if their gender is best described as “cisgender male”, “cisgender female”, or “transgender or gender non-conforming (GNC)”. (For clarification on the terms cisgender and transgender, you may refer to [this description](#).)

Gender identity	Floor worker	Supervisor
Trans or GNC	1994	6
Cis woman	40354	8809
Cis man	43372	5471

The table above summarizes their findings based on this alternative classification of gender identities. Determine whether or not there is evidence of discrimination by testing the claim that the distribution of gender identities across each level of workers is the same. Use an $\alpha = 0.01$ level of significance. Show all your work and interpret your conclusion within the context of the problem and the appropriateness of the required assumptions.

¹The data for these problems is hypothetical.

School segregation

For problems 3-5², consider the following setting.

In 2000 a particular school district had a total of 7982 students enrolled in school. At this time, the percent of white students who were exposed to school poverty³ was 30% whereas this was 34.5% for Black students. The number of white students in this district in 2000 was 5184 and the number of Black students was 1035. (Note these are the demographics for students who identify as one of these two races alone.)

In 2015, the number of students enrolled in school in this district increased to 11977. The percent of white students exposed to school poverty increased to 47.8% and the percent for Black students increased to 76.1%. The number of white students in this district in 2015 was 7306 and the number of Black students was 2037.

Problem 3

Suppose we want to determine if, relative to the increased size of the school district from 2000 to 2015, is the difference between the proportion of white students exposed to poverty likely due to something besides chance. Perform an appropriate statistical hypothesis test at an $\alpha = 0.05$ confidence level to answer this question. Show all your work and interpret your conclusion within the context of the problem and the appropriateness of the required assumptions.

Problem 4

Perform the same analysis as you did in Problem 3 but this time to determine if there is statistical evidence of an actual difference between the proportion of Black students exposed to poverty.

Problem 5

Now suppose we want to estimate the difference in the proportion of students who are exposed to school poverty along racial categories. Find a 95% confidence interval for the difference between the proportion of Black students exposed to school poverty and the proportion of white students exposed to school poverty in the year 2015. Make sure you specify the point estimate, the margin of error, and the estimate of the standard deviation of the difference in sample proportions. Then interpret your interval in the context of the problem.

²The data for these problems is also hypothetical but are informed by the summary statistics the 2020 paper *Racial Segregation and School Poverty in the United States, 1999–2016* by E. Fahle, S. Reardon, D. Kalogrides, E. Weathers, and H. Jang.

³Exposure to school poverty was determined by whether or not a student attended a school where more than 75% of students were eligible for the Federal free and reduced-price lunch program.