PSC 103B

Homework 6

Winter 2024

**Instructions**

Please use R/RStudio to complete the following questions. You will submit your filled-out version of this document **as a PDF** on Canvas. Make sure your PDF looks as expected before submitting. Unless otherwise specified, please **always** **include the code you used to generate your answer for each question, or the steps you used to calculate the answer (when relevant), as well as the final answer and/or relevant output** (output is what comes out in the console when you run an R command, e.g., the results of the model when you run summary(fit1)).It’s a good idea to organize your R code in the R script and save it, so if you need to modify or recalculate one of the questions, that’s easy to do (see the R scripts provided in previous weeks for tips on how to organize your code). If you copy and paste code or output into this document (screenshots are also acceptable), **please format the code and output using a fixed-width font** (e.g., Courier) so it’s easier to read.

You may consult with your classmates while working on the assignment, but **you must do all the work yourself – everything you turn in must be your own code and words**. Academic dishonesty will not be tolerated.

Please submit **a pdf version of this document** with your answers on Canvas by **1:59pm on Friday, March 12.**

**Question 1**

We’ll be using the observed frequencies that we used in lab of Fall 2022 enrollment across the different colleges of UC Davis. I have repeated this information below:

|  |  |
| --- | --- |
| **College** | **Observed Frequency** |
| College of Letters & Sciences | 417 |
| College of Agricultural and Environmental Sciences | 223 |
| College of Biological Sciences | 216 |
| College of Engineering | 144 |
| Total | 1000 |

Suppose we were interested in conducting a goodness-of-fit test to determine whether enrollment in these colleges matches the proportions observed in 1993. Here are the frequencies that were observed in 1993.

|  |  |
| --- | --- |
| **College** | **Observed Frequency** |
| College of Letters & Sciences | 465 |
| College of Agricultural and Environmental Sciences | 223 |
| College of Biological Sciences | 191 |
| College of Engineering | 121 |
| Total | 1000 |

What proportion of students were enrolled in each college in 1993? (1 point)

**Question 2**

Use the probabilities of Question 1 to write the null and alternative hypotheses for the Goodness-of-Fit test (2 points).

**Question 3**

What are the expected frequencies? (1 point)

**Question 4**

Conduct the chi-square goodness-of-fit test in R. Show your code and output (1 points).

**Question 5**

Do you reject or fail to reject the null hypothesis? Does 2022 enrollment in the different colleges match the proportions of 1993? (1 points)

**Question 6**

Now suppose we were interested in whether students’ choice of college was related to how long it took them to graduate. The frequencies from the 2015 cohort (my entry year!) are provided in a table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **CLAS** | **CA&ES** | **CBS** | **COE** | **Total** |
| **Graduated at 3 years or less** | 72 | 30 | 28 | 5 | 135 |
| **Graduated at 4 years** | 1238 | 677 | 803 | 444 | 3162 |
| **Graduated at 5 years** | 378 | 216 | 329 | 243 | 1166 |
| **Graduated at 6 years** | 51 | 43 | 38 | 482 | 614 |
| **Total** | 1739 | 966 | 1198 | 1174 | 5077 |

What are the null and alternative hypotheses for the test of independence? (2 points)

**Question 7**

Conduct the chi-square test of independence. Show your code and output (1 points).

**Question 8**

Do you reject or fail to reject the null hypothesis? What does this lead you to conclude about time to graduation and college? (1 points)