Tasks

For this week's problem set, you are tasked with helping Dr. Mark Gibney. As a bit of background information, Dr. Gibney has been preoccupied of late with something called *grade inflation*. Grade inflation is typically understood as the devaluing of grades given to students in an educational setting. For example, the grade of A may have been reserved only for students displaying exceptional mastery of the course material in 1973 but in 2021, students displaying a *merely* good understanding of the course material might also be receiving the grade of A. In other words, grade inflation exists when the meaning of a grade is degraded. Rather than signifying truly remarkable and outstanding performance, an A today might simply mean good.

If overtime more and more A's or B's are received by students at a university relative to C's, D's, and F's, or if the average student GPA increases overtime, this could be considered evidence of grade inflation at that university. Of course increasing grades and GPA's over time could be explained by other factors. It could be that the meaning of grades has changed and that instructors are becoming easier graders, but it could also be the case that the *quality* or *ability* of students has improved. Grades are improving not because a lowering of standards but because of the recruitment of better students to the university. Similarly, the quality of instruction surely would be correlated with grades as better instruction leads to improvement in student performance and hence grades. So it could be that grades improve because the University has hired better instructors.

Whatever the case may be, I have gotten my hands on grading records at UNCA for the last 20 years. These data are contained in the accompanying file `ProblemSet3.Rdata`. The file contains two data-frames – `GradingScale` and `Grades`. The former just contains two variables, the letter grade and the numeric grade points associated with each letter grade (e.g., an A is worth 4 grade points, an F zero). The second data-frame contains 13 variables. The first variable (i.e., vector) is `Department`. It is a character vector containing the names of the academic departments at UNCA. The second variable is `Year` and it contains a year between 2001 and 2020. Combined these two variables constitute the *Unit of Analysis* – which is to say each row or record in our data-frame contains data for a specific department for a specific year. Therefore the units we are analyzing are *department-years*. For example row 3 of the data-frame contain information for the Africana Studies Department in 2003 and row 631 contains data for Women, Gender & Sexuality Studies in 2015. The remaining variables or vectors (`A`, `A.Minus`, `B.Plus`, & cetera) contain counts of grades. For example, the vector `B.Minus` contains the number of B– grades that were assigned by a given department in a given year. Africana Studies, for instance gave out 17 A's, 3 B's, and 3 F's in 2006 across all classes taught in that year in that department.

On to the tasks ...

1. How many units or records or rows of information does the `Grades` data-frame have?
2. Replace the character vector `Department` in the `Grades` data-frame with the factor version and determine how many records (i.e., rows) we have for each department? Is the dataset *balanced* in the sense that for each department we have an equal number of records?
3. Over the entire 20 year period, how many grades were assigned at UNCA?
4. Over the entire 20 year period, how many grades of A, A–, B+, etc were assigned at UNCA?
5. Over the entire 20 year period, what percentage of all grades at UNCA were A's A–'s, B+'s, etc?

6. Over the entire 20 year period, how many grades were assigned in the Political Science Department?

7. Over the entire 20 year period, how many grades of A, A–, B+, etc were assigned in the Political Science Department?

8. Over the entire 20 year period, what percentage of all grades in the Political Science Department were A's A–'s, B+'s, etc?

9. Obviously, larger department's such as Psychology will assign more grades in a given year than smaller departments such as WGSS. Additionally over time more students might enroll at UNCA leading to more grades being assigned in some years compared to others. To account for this and allow for the comparison of departments and to identify trends over time, computing the Grade Point Average (GPA) per department and/or year would be helpful.[1]

   (a) Calculate the GPA for all of UNCA for each year (i.e., compute 20 GPAs, one for each year) and store them in a vector called `UNCA_GPA`.

   (b) Calculate the GPA for the Political Science Department for each year (i.e., compute 20 GPAs, one for each year) and store them in a vector called `POLS_GPA`.

   (c) Create a data-frame called `GPA` that contains the two vectors you created for the tasks above (i.e., (a) and (b)) and a variable for the year (i.e., that year vector should be a sequence from 2001 to 2020). In total the data-frame should have 20 rows and 3 columns.

   (d) In a short paragraph and based on your findings above, discuss if there is any evidence of grade inflation at UNCA.

Extra-Credit

Explore the `Grades` data-frame further and report at least one additional finding or statistic or interesting pattern that may be of interest to Dr. Gibney.

Note: *The above tasks might appear tedious and may require quite a bit of duplication and many lines of code. We will learn how to make code more efficient in the next weeks (e.g., via loops, and vectorization). For now just slog through it. All problems can be solved by relying on functions and commands we have covered in class. In your code, do not use functions or commands we have not covered in class.*

[1] You may compute GPA similar to how you would compute GPA for a single student. Let's pretend here that all courses at UNCA have the same number of credit hours. For example: A student who took three classes in a year and received an A, a B– and a D+, would have a GPA of $\frac{4.00+2.66+1.33}{3} = 2.66$.