14 - Assignment (due November 2)

Problem 1A (comprehension)

What's the matrix equation (including the design matrix) for the linear model built from the dataset below where the column "Salary" serves as the response variable and all other columns serve as predictor variables?

Employee_ID	Employment Status	College	Years_Experience	Salary (Y)
1	Full-time	Yes	2	52,000
2	Part-time	Yes	5	68,000
3	Full-time	No	3	115,000
4	Full-time	No	7	85,000

Assume that "Employment Status" has two levels: full-time vs part-time.
Assume that "College" has two levels: yes or no.

Problem 1B

The above dataset turns out to come from a larger dataset with the same number of columns but to actually have 2000 rows.

Suppose β_1 is the coefficient of "Employee_ID" in the best-fitting linear model for this larger dataset. What do you expect to be the value of β_1? Justify your answer.

Problem 2 (applied)

For this problem, we use the following dataset:
https://unsw-risk-and-actuarial-studies.github.io/ACTL3142/M1-Introduction/College.csv

Problem 2A

Write SAS code to build a linear model where "Private", "Apps", "Accept", "Enroll", "S.F.Ratio", "Expend" are the explanatory variables used to predict "Grad.Rate".

Write down the equation of this linear model if possible or explain why this is not possible to do.

If there is multicollinearity, explain which subset of variables exhibit this and justify your answer by computing the R^2-value for this subset of variables.

Problem 2B

Write SAS code to build a linear model where "Private", "Enroll", "S.F.Ratio", "Expend" are the x's used to predict "Grad.Rate". Then write down the equation of this linear model if possible or explain why this is not possible to do.

Problem 2C

From the linear model in Problem 2B, simplify your model when it is only applied to private schools.

Problem 2D

Find and interpret a 95% confidence interval for the "Expend" coefficient in the linear model you constructed in **Problem 2B**.

Problem 2C

Given all variables in the model in **Problem 2B**, which variables are significant predictors of kid’s cognitive test score?