Statistical Methods for Data Science
Mini Project 5

Instructions:

- **Due date: Wednesday, April 20.**
- **Total points = 20**
- Submit a typed report and include any relevant plots and send it to Lei Zhang (lxz096120@utdallas.edu).
- You must use the following template for your report:

  Mini Project #
  Name

  Provide the R codes in an appendix. Your code must be annotated. No points may be given if a brief look at the code does not tell us what it is doing.

For all problems, write the appropriate hypotheses, assumptions (if needed), null distribution of the test statistics and also the expression of the P-value along with proper conclusion. Points will be deducted if the above are not mentioned.

Exercise 1 (7 points):
The following table shows the Myers-Briggs personality preferences for a random sample of 389 past computer science graduates (from a University) in the listed professions

Occupation	Personality preference Type	
	E	I
Faculty	62	45
Data Scientist	56	81
Entrepreneur	94	51

Determine if there is any association between the listed occupations and the personality preferences at 5% level of significance.

Exercise 2 (6 points):
Makers of generic drugs are required by the FDA to show that the extent of absorption of their drug in blood does not differ significantly from the "brand-name" drug that they imitate. To show this, 20 healthy nonsmoking male subjects are selected. For each subject, one of the two drugs is randomly chosen and given first. Then after a washout period, the other drug is given. In both the cases, the absorption of the drug in the blood is measured. The dataset stored in the file medicine.txt gives the measurements taken on 20 subjects. Do the drugs differ significantly in absorption?

Exercise 3 (7 points):
This landmark experiment in genetics investigated whether, for a certain kind of sweet pea plant, the traits “flower color" and “pollen grain type" are inherited independently or not. Flower color can be purple (P) or red (R), but the purple color is dominant) and Grain type can be Long (L) or Round (R), but long grain type is dominant. According to Mendel's law of independent segregation, the genes for these two traits segregate independently and the “flower color" and “pollen grain type" combinations-P&L, P&R, R&L and R&R are expected in 9:3:3:1 ratio. The following are the observed frequencies for each combination when the experiment was carried out on 256 sweet pea plant.

“flower color" and “pollen grain type" combinations	Observed Frequencies
P&L	177
P&R	15
R&L	15
R&R	49

Write your conclusion at 5% level of significance.