

Submit an R notebook with comments, code and results.

**Problem 1 (10 points):**

We would like to study the correlation between mental and physical well-being. An analysis of birthdays and death days of famous people could be used as further evidence in the study of this correlation. To use these data, we are supposing that being able to look forward to something betters a person's mental state; and that a famous person would probably look forward to his or her birthday because of the resulting attention, affection, and so on. If a famous person is in poor health and dying, then perhaps anticipating his birthday would "cheer him up and therefore improve his health and possibly decrease the chance that he will die shortly before his birthday." The data might therefore reveal that a famous person is less likely to die in the months before his or her birthday and more likely to die in the months afterward.

Months before (-) / After (+) Birthday	Number of Deaths
-6	90
-5	100
-4	87
-3	96
-2	101
-1	86
0	119
1	118
2	121
3	114
4	113
5	106

1. (2 point): Specify an appropriate null hypothesis and the distribution you would compare against. (max 5 sentences)
2. (3 points): Specify the relevant statistic, degrees of freedom and show your calculation (max 5 sentences)
3. (1 points): Calculate the p-value (max 5 sentences)
4. (2 points): What would you conclude for 10% and 10% level of significance. Provide detailed comment (max 10 sentences)
5. (2 points): Submit your work before deadline.

**Problem 2 (50 points):**

An experiment was devised to test the effects of running 3 different types of gasoline with 3 possible types of additive. The experiment called for 9 identical motors to be run with 5 gallons for each of the pairs of gasoline and additives. The following data resulted.

Gasoline	Additive 1	Additive 2	Additive 3
1	124.1	131.5	127
2	126.4	130.6	128.4
3	127.2	132.7	125.6

- a) (5 points): Test the hypothesis that the gasoline used does not affect the mileage.

- b) (5 points): Test the hypothesis that the additives are equivalent.
- c) (5 points): What assumptions are you making?

To allow for the possibility of an interaction effect between gasoline and additive, it was decided to run 36 motors — 4 in each grouping. The following data resulted.

Gasoline	Additive 1	Additive 2	Additive 3
1	124.1	131.5	127
2	126.4	130.6	128.4
3	127.2	132.7	125.6
1	124.5	131.2	126.8
2	126.1	130.2	128.5
3	127.0	133.1	125.7
1	123.9	131.2	127.1
2	126.4	130.4	128.2
3	127.2	133.3	125.3
1	124.4	131.1	126.6
2	126.2	130.1	128.4
3	127.1	133.2	125.5

- a) (3 points): Specify the average value for each gasoline.
- b) (3 points): Specify the average value for each additive.
- c) (4 points): Specify the gasoline and additive combination that produced maximum value and minimum value on average
- d) (5 points): Do the data indicate an interaction effect? Justify your answers with numerical calculations and explain the results (max 5 sentences)
- e) (5 points): Do the gasolines appear to give equal results? Justify your answers with numerical calculations and explain the results (max 5 sentences)
- f) (5 points): Test whether or not there is an additive effect or whether all additives work equally well. Justify your answers with numerical calculations and explain the results (max 5 sentences)
- g) (5 points): Specify your final model
- h) (5 points): You are tasked to choose the gasoline and additive combination that will produce the maximum value. Which one will you choose. Justify your answer.

### Problem 3 (45 points):

In our local town people are affiliated with one of 3 parties, A, B and C. We would like to study if there is any correlation between a person's gender and political affiliation.

In the first part you are supposed to collect the data. Here we will not ask to collect the data explicitly, but require you to specify the data collection process, so you can make better conclusion:

- a) (10 points): Specify how would you go about collecting the data. Make sure to provide what data you want to collect, any format, the key factors you want to keep in consideration, sample size and so on (max 15 sentences)

Now that you have thought about the experiment and discussed the process of data collection with your colleagues, they will help you in getting the actual data collected so you can do the analysis.

The table below shows the data collected by your colleagues and summarized for your analysis:

Gender / Party →	Party A	Party B	Party C
Women	68	56	32
Men	52	72	20

Now let us start the analysis:

- b) (2 points): How many people were sampled in total?
- c) (2 points): How many men and women were sampled?
- d) (2 points): Which party has the greatest number of representatives?
- e) (2 points): State your null hypothesis.

Based on the null hypothesis and the observed data, answer the following questions:

- f) (6 points): How many females are expected to be in party A, B, C respectively?
- g) (6 points): How many males are expected to be in party A, B, C respectively?
- h) (2 points): What is the test statistic you would use? Justify (max 3 sentences)
- i) (8 points): Calculate the value of the test statistic based on the number you calculated above and the data you collected. Do not use any R function that directly calculates the test statistic. Show your work explicitly.
- j) (5 points): What would you conclude based on the test statistic? Justify (max 5 sentences)