**Submit an R notebook with comments, code and results and discussions**

**Problem 1 (15 points):**

In a test of the ability of a certain polymer to remove toxic wastes from water, experiments were conducted at three different temperatures. The data below give the percentages of the impurities that were removed by the polymer in 21 independent attempts.

| Low | Medium | High |
|---|---|---|
| 42 | 36 | 33 |
| 41 | 35 | 44 |
| 37 | 32 | 40 |
| 29 | 38 | 36 |
| 35 | 39 | 44 |
| 40 | 42 | 37 |
| 32 | 34 | 45 |

a) (3 point): Specify an appropriate null hypothesis (max 5 sentences)
b) (5 point): Test the hypothesis that the polymer performs equally well at all three temperatures at 5 percent level of significance
c) (5 point): Test the hypothesis that the polymer performs equally well at all three temperatures at 1 percent level of significance
d) (2 point): State your conclusion from the analysis.

**Problem 2 (15 points):**

An emergency room physician wanted to know whether there were any differences in the amount of time it takes for three different inhaled steroids to clear a mild asthmatic attack. Over a period of weeks she randomly administered these steroids to asthma sufferers, and noted the time it took for the patients' lungs to become clear. Afterward, she discovered that 12 patients had been treated with each type of steroid, with the following sample means (in minutes) and sample variances

| Steroid | Mean | Variance |
|---|---|---|
| A | 32 | 33 |
| B | 40 | 44 |
| C | 30 | 40 |

a) (5 point): Test the hypothesis that the mean time to clear a mild asthmatic attack is the same for all three steroids. Use the 5 percent level of significance.
b) (10 point): Find confidence intervals for all differences of means ( ) that, with 95 percent confidence, are valid.

**Problem 3 (30 points):**

You are budding gardener and would like to grow snow peas (because you love them) in your garden. You have a lot of space in your garden and there are different regions that you can grow: Full Sun (FS), Partial Shade (PS), Shade (SH). You want to take a data driven approach to choose the best spot to plant a lot of them to maximize your return. You have planted some seeds in each of these areas. After a few weeks, finally you are starting to see some peas on the plants, and you collect the amount of peas harvested from each area. This is shown in the table below:

| Yield | Area |
|---|---|
| 18.6 | FS |
| 19.9 | FS |
| 17.1 | FS |
| 18.4 | FS |
| 17.8 | FS |
| 19.9 | FS |
| 16.8 | FS |
| 16.5 | FS |
| 18.6 | FS |
| 16.4 | FS |
| 17.8 | SH |
| 18.1 | SH |
| 18 | SH |
| 16 | SH |
| 16.7 | SH |
| 17.7 | SH |
| 16.7 | SH |
| 15.1 | SH |
| 17.6 | SH |
| 17.2 | SH |
| 17.8 | PS |
| 18.8 | PS |
| 21.4 | PS |
| 19.1 | PS |
| 21.1 | PS |
| 17.9 | PS |
| 19 | PS |
| 19.2 | PS |
| 16.8 | PS |
| 17.8 | PS |

a) (5 points): State your null hypothesis. (Max 5 sentences)
b) (5 points): What analysis strategy (you have learned so far would you choose to analyze) (max 5 sentences)
c) (5 points): Calculate the test statistic and p-value.
d) (5 points): Calculate the residual error.
e) (5 points): State your conclusion.

You are not fully happy with your analysis, and you think you might be able to make a better decision if you reduce the residual error.

f) (5 points): What approach can you take to reduce the residual error. Discuss at least 2 ideas and justify (max 10 sentences)

**Problem 4 (40 points):**

You remember that you have also noted down the height of the plants when you were harvesting the snow peas. Maybe you can use this covariate to reduce the residual error from problem 2 and have more confidence in your decision (after all you are going to make a big decision). The data with the height of the plants are shown below.

| Yield | Area | Height(inches) |
|---|---|---|
| 18.6 | FS | 10 |
| 19.9 | FS | 11.5 |
| 17.1 | FS | 13 |
| 18.4 | FS | 13.2 |
| 17.8 | FS | 14.2 |
| 19.9 | FS | 13.3 |
| 16.8 | FS | 11.9 |
| 16.5 | FS | 12.5 |
| 18.6 | FS | 10.5 |
| 16.4 | FS | 11.2 |
| 17.8 | SH | 8.2 |
| 18.1 | SH | 9.1 |
| 18 | SH | 7.2 |
| 16 | SH | 7.9 |
| 16.7 | SH | 8.3 |
| 17.7 | SH | 9.1 |
| 16.7 | SH | 8.5 |
| 15.1 | SH | 8.7 |
| 17.6 | SH | 9.4 |
| 17.2 | SH | 9.6 |
| 17.8 | PS | 12.5 |
| 18.8 | PS | 13.2 |
| 21.4 | PS | 14.1 |
| 19.1 | PS | 13.2 |
| 21.1 | PS | 14.5 |
| 17.9 | PS | 13.3 |
| 19 | PS | 12.9 |
| 19.2 | PS | 12.5 |
| 16.8 | PS | 11.7 |
| 17.8 | PS | 13.3 |

a) (5 points): What analysis strategy would you use to take advantage of the additional data you have (you have learned so far would you choose to analyze) (max 5 sentences)
b) (10 points): What is the dependence of the yield on the height for each of the areas? Does it differ significantly between the 3 areas. What can you conclude (max 10 sentences)
c) (5 points): Calculate the test statistic and p-value.
d) (10 points): Calculate the residual error and compare that to what you found in problem 2. Discuss your findings (max 10 sentences)
e) (5 points): Specify your final model.
f) (5 points): State your conclusion (max 5 sentences)