Session 9: hypothesis testing (Lab Report 3)

Background

The North Carolina Department of Environmental Quality (DEQ) has observed neonicotinoid insecticide concentrations in groundwater across a few local potato farming communities (*note that this is a hypothetical scenario*). It is believed that the neonicotinoid insectides are leaching into the groundwater through their application to potato fields. NC DEQ has hired you as a consultant to evaluate how neonicotinoid insecticide leaching could be mitigated. Specifically, they are interested in knowing whether the active ingredient and/or application style of different neonicotinoid insecticides influences the magnitude of leaching. NC DEQ plans to offer incentives to potato growers who adopt the use of a neonicotinoid insecticide with the active ingredient and application method that minimizes groundwater impacts. They will determine which insecticide to promote per your recommendations.

To develop your recommendation to NC DEQ, you analyze data that were collected in 2011 and 2012 as part of a study led by Dr. Anders Huseth (Entomology and Plant Pathology, NCSU) and Dr. Russell Groves (Entomology, UW-Madison). In this study, researchers evaluated how neonicotinoid (1) application strategy and (2) active ingredient influenced groundwater quality. Potato crops received each of the four neonicotinoid applications shown in Figure 1 (foliar spray, seed treatment, in-furrow spray, impregnated gel) with one of two active ingredients (imidacloprid or thiamethoxam). Groundwater samples were collected throughout the growing season. Additionally, groundwater samples were collected in potato fields where **no insecticide** was applied ("untreated control").

Your primary goal is to determine the neonicotinoid insecticide products (active ingredient + application method) for which there is **no evidence of neonicotinoid leaching to groundwater.**

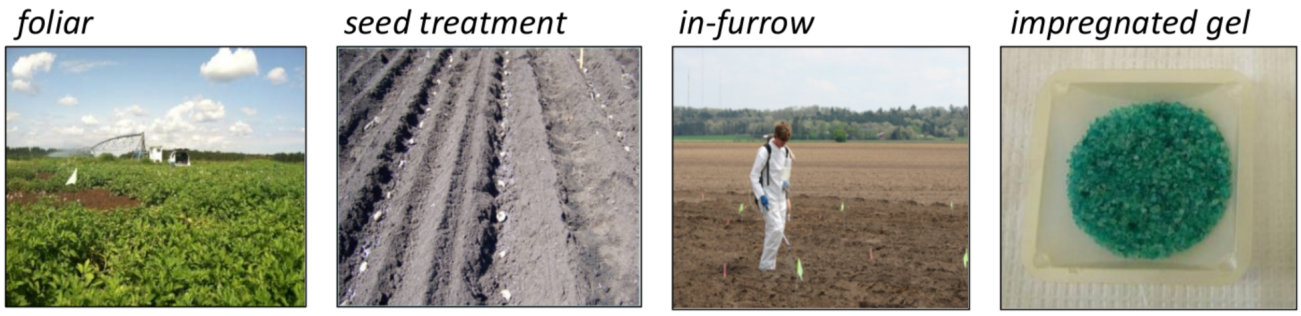


Figure 1: Neonicotinoid insecticide application methods. Credit: Anders Huseth.

Data

The data are included in two files: "imidacloprid.csv" and "thiamethoxam.csv". These files include the measured groundwater concentrations (ug/L) of imidacloprid and thiamethoxam, respectively.

Analysis objectives

All analysis should be performed in R. It is suggested that these objectives be addressed in the order they are listed.

1. Determine whether there is a statistically significant difference (rejection level of 0.05) between the application methods of different ingredients relative to the untreated control using a **one-tailed paired t-test**. You should conduct 8 paired t-tests total (4 application methds x 2 active ingredients).

- Use the R function `t.test()` to apply the paired t-tests. Specifically, you will have to run this function with the following arguments: `t.test(*APPLICATION/INGREDIENT*, *UNTREATED CONTROL*, alternative = "greater", paired = TRUE)`. Read the help documentation to determine what these arguments mean. Note that the alternative hypothesis has been specified as an argument - be sure to report the null and alternative hypotheses in your report.

2. Confirm the t-test results from Objective 1 by calculating the mean and standard deviations of the differences (differences being between the concentrations of each application/ingredient from the untreated control), number of observations (n), and corresponding t-statistics and p-values for each of the 8 tests. Report these values (mean, standard deviation, n, t-statistic, p-value) for each application/ingredient in one table in your report. These values all need to be in one dataframe in order to perform Objective 3 - this new dataframe should have 1 column for the means, 1 for standard errors, 1 for n, 1 for t-statistics, and 1 for p-values.

- Use the following code to create this new data frame:

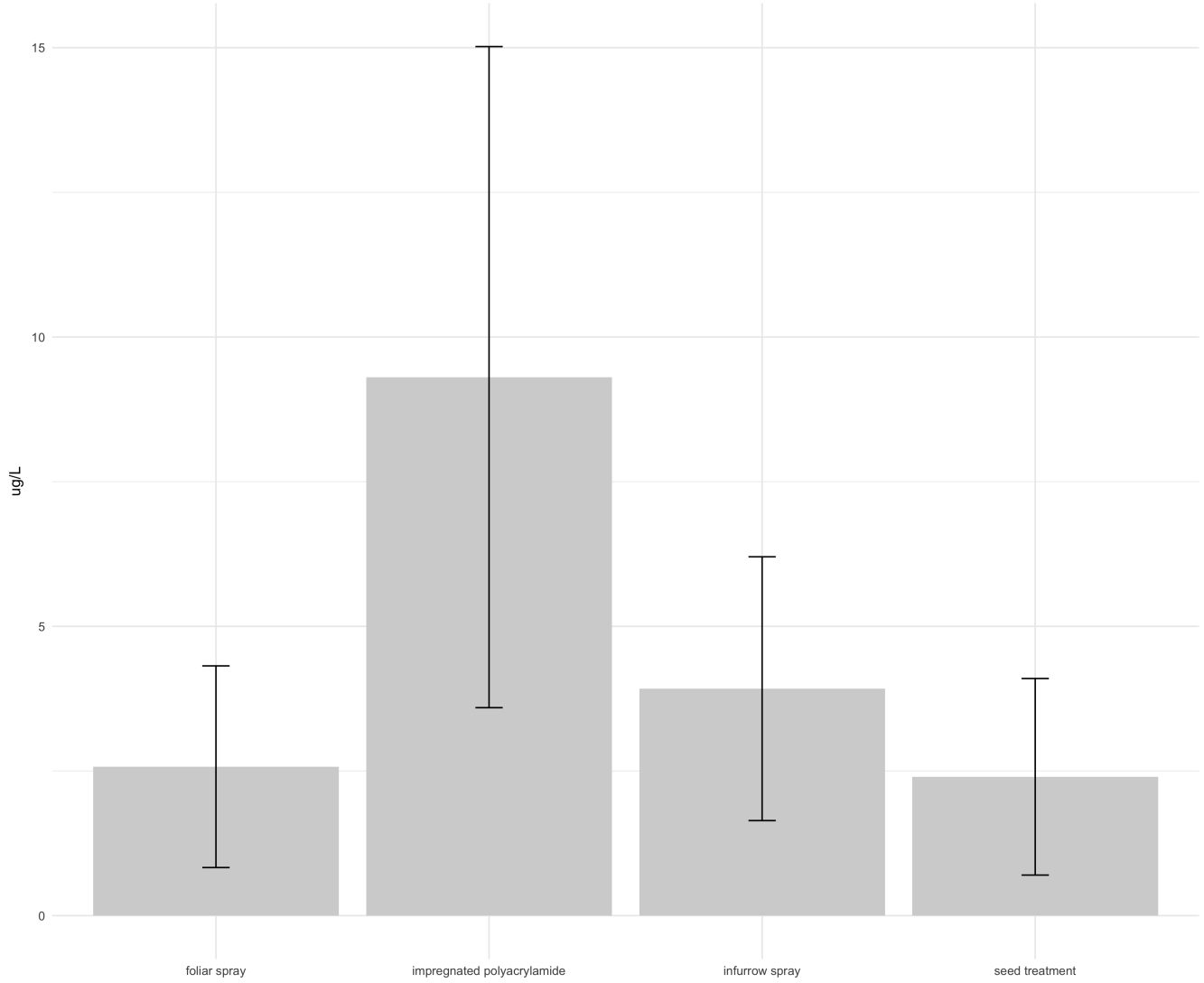
```
d %>% transmute(`foliar spray`=`foliar spray`-`untreated control`,
            `impregnated polyacrylamide`=`impregnated polyacrylamide`-`untreated control`,
            `infurrow spray` = `infurrow spray`-`untreated control`,
            `seed treatment` = `seed treatment`-`untreated control`) %>%
  gather(`foliar spray`,`impregnated polyacrylamide`,`infurrow spray`,`seed treatment`, key = "del",valu
  group_by(del) %>%
  summarise(Mean = , SD = , n = n()) %>%
  mutate(t = ) %>%
  mutate(p = )
```

- To calculate the p-value, use the `pt()` function. Note that this is one of the R distribution functions (e.g., pnorm, plnorm, pgamma, etc.). `pt()` calculates the left-tail cumulative probability. You already calculated the p-values in Objective 1, so you can use your findings from Objective 1 to validate the results you calculate for this objective.

3. For each active ingredient, create a bar plot of the **mean differences from the control** (the values in the "mean" column in the dataframe you created for Objective 2) with error bars that correspond to 90% confidence intervals. One of the two plots you need to create has been shown below as an example (note that you will not receive credit if you copy and paste the plot into your report).

- Include the lines of code below to create your ggplot. In your report, **explain where the 1.796 values came from**.

```
  geom_col(fill = "lightgrey")+
  geom_errorbar(aes(ymin = Mean-1.796*SD/sqrt(n), ymax = Mean+1.796*SD/sqrt(n)), width = 0.1)
  # "width" corresponds to the width of the lines that cap the error bars; ymin and ymax correspond to
```

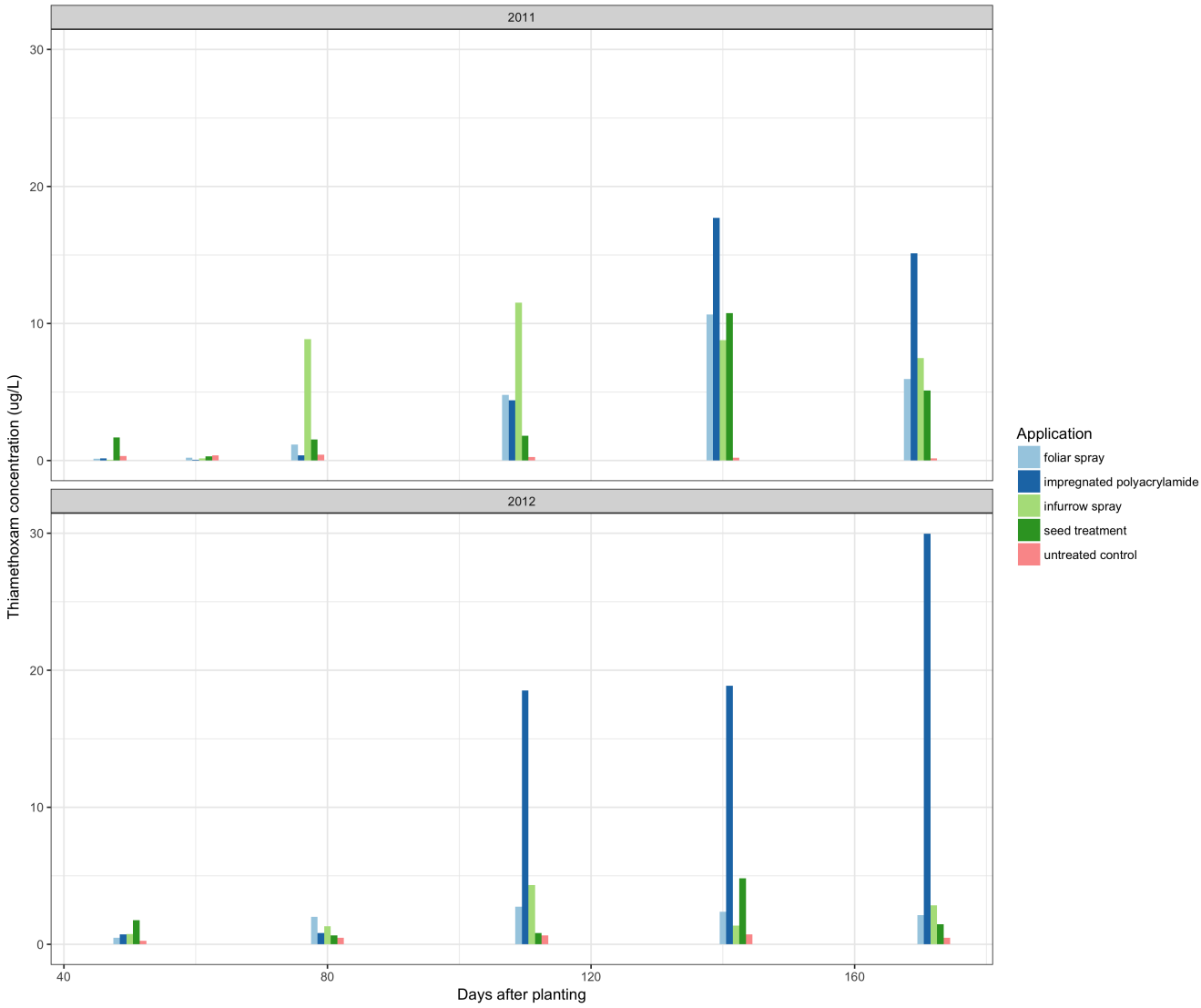


Mean differences in groundwater thiamethoxam concentrations between insecticide application methods and the untreated control. Error bars correspond to a 90% confidence interval.

4. For each active ingredient, use bar plots to show the groundwater concentrations associated with each application method. One of the two plots has been provided to you below as an example.

- Include the lines of code below to create your ggplot.

```
geom_col(position = "dodge", width = 5)+
scale_fill_brewer(palette = "Paired")+
facet_wrap(~year, ncol = 1)
```



Groundwater thiamethoxam concentrations among insecticide application methods and the untreated control as a function of days after planting; measurements from the 2011 study are plotted in the upper half of the plot, measurements from the 2012 study are plotted in the lower half of the plot.

Report

The primary objective of your report is to summarize study, detail your methods, and offer recommendations to NC DEQ regarding the most suitable insecticide to incentive use of among potato growers. Although you can work together to perform the analysis, the report **must be in your own words** and all material should be original. Any reports with evidence of plagiarism will receive no credit.

Be sure to address each of the analysis objectives in your report. Your report should be structured as follows:

- Title page - title (create your own title), name, date
- Executive abstract - summarize the report in 4-6 sentences; be sure to include your main results and recommendations, as this is the most important information included in abstracts.

- Introduction - provide all background information needed to understand the contents of the report. Be sure to *paraphrase* the writeup included here; do not copy directly.
- Methods - explain the statistical methods you used, both based on the theory and the specific application at hand. Demonstrate that you understand how and why you used the statistical principles in this case study. Include equations when relevant. **You do not have to detail the steps you took in R.**
- Results - state the results objectively. Explain what the figures and numbers you've produced from this analysis are telling you.
- Recommendations - explain to NC DEQ what evidence there is that certain neonicotinoid insecticide products (active ingredient/application method) offer protections against insecticide leaching to groundwater.
- Appendices - include your R code (should be neatly organized and professional).

Additional notes on formatting:

- All figures and tables should be captioned, numbered, and referred to in the text.
- Table captions are placed *above* the table. Figure captions are placed *below* the figure.
- Equations should be centered and included on their own lines.