**Analyze the data sets below and write a brief report that includes a summary of findings, a graphical display and a section describing the methods used to answer the questions of interest.**

1. The data in the file wine.csv (in the datasets folder on Canvas) give the average wine consumption rates (in liters per person) and number of ischemic heart attack deaths (per 1000 men aged 55 to 64 years) for 18 industrialized countries.

Do these data suggest that heart disease death rates are associated with average wine consumption? If so, how can that be described?

Do any countries have substantially higher or lower death rates than others with similar wine consumption rates?

Analyze the data and write a brief report that includes a summary of findings, a graphical display and a section describing the methods used to answer the questions of interest.

1. Meadowfoam is a small plant that grows in Pacific Northwest and is domesticated for its seed oil. A study was set up to determine if meadowfoam can be made into a profitable crop. In a controlled growth chamber, the plant was grown at 6 different light intensities and two different timings of onset of light treatment. The outcome of interest is the number of flowers per plant which was measured by averaging numbers of flowers produced by 10 seedlings in each group. Growth was replicated at each combination of time and light intensity.

a. First put the data into a dataset with four variables: number of flowers, light intenisty, timing and replicate.

b. Create a categorical form of the light intensity with 6 categories.

c. The research questions are: **What are effects of intensity and timing? Is there an interaction between the two factors?**

d. First create an analysis of variance using timing and the categorical form of the light intensity variable. Determine if there is an effect of each factor.

e. Then create an interaction between light intensity and timing by multiplying the two variables and test for the presence of an interaction.

f. Now repeat the process but using light intensity as a continuous variable.

g. Then perform F-tests to compare the four model you have created (light as continuous and categorical with and without the interaction)

h. Predict the number of flowers grown at each combination of light and timing for each of the four models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Timing | Light | Model 1 | Model 2 | Model 3 | Model 4 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

i. Compare each prediction to the observed number of flowers and calculate the difference (observed – predicted). This is the residual. Calculate the residual mean squared error for each model by adding the squared residuals together and dividing by the number of residual degrees of freedom. This should equal the mean squared error in each ANOVA table.

j. Now plot the residuals vs. the predicted for each model and see if there are any patterns. If you see any, what might you do to remove them?

k. Finally, take the model you think describes the data the best and write a short report for your grandmother who would like to grow these flowers carefully explaining to her how she should best grow them and why. Note that your grandmother is curious about how much changes in light and timing might affect her flowers and how sensitive her results will be to the settings she makes.