Problem 1: An MU scientist is interested in determining how well three different types of geranium grow in different amounts of direct sunlight. To perform the experiment, she obtains space in a greenhouse and 18 plants from each of the three types of geranium. She then sets up “tents” to keep some of the plants from receiving total sunlight. She makes 2 mesh tents (partial sunlight) and 2 canvas tents (no direct sunlight). She then places three plants of each of the types into each of the tents. She also places three of each type of plant in two other locations to receive full sunlight. After three weeks she measures the amount of growth for each plant.

(7 pts.) Write down the model for this design. Make sure to mention which of the observations you expect to be correlated. Also, mention which terms in your model are fixed and which are random.

\[ Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijkl} \]

\[ \epsilon \sim N(0, \sigma^2 \Sigma) \]

where observations in the same tent are correlated.

(4 pts.) Look at the output provided for this problem. Which of the fixed effects are significant? What can you conclude from this information?

The interaction between amount of sun and type of plant is significant. This means that different types of plant grow better in different amounts of light.

(8 pts.) What is the estimate for the error variance? How about the correlation between observations? What can you conclude about these two types of errors?

\[ \hat{\sigma}^2_{error} = 287.08 \]

The correlation between observations in the same tent is 74.7248/287.08 = .2605.

The amount of random error is large relative to the correlation between observations in the same tent.

(6 pts.) Suppose that you have a partially sunny spot where you desire to grow geraniums. Which of the three types do you believe will grow best in your yard?

Type B geraniums have the largest predicted growth in these conditions (94.05) as compared to type A (62.45) and type C (50.38).
Problem 2: Consider the output for problem 1, and answer the following questions.

(7 pts.) Define the AIC criteria. Make sure to mention what the different values in your formula are. Calculate the value of the AIC for the output in problem 1. Note that your answer may differ from what SAS would get due to different definitions.

\[ AIC = \ell(\hat{\theta}, \hat{\beta}) - q. \]

Here \( \ell \) is the log-likelihood and \( q \) is the number of covariance parameters.

\[-2\ell(\hat{\theta}) = 402.1 \iff \ell = -201.05 \quad q=2\]

\[ AIC = -201.05 - 2 = -203.05.\]

\[ (AIC_{(SAS)} = -2AIC = -2(-203.05) = 406.10).\]

(7 pts.) Predict the average response for the "B" type of geranium grown in full sunlight.

See the last part of question 1. (Similar)

\[ \hat{\gamma} = 50.3833 + 51.6500 + 43.6667 - 62.1500 = 83.55. \]

(7 pts.) Perform a likelihood ratio test to compare the model indicated on this output to a model which does not include an interaction between the different types of geranium and the amount of shade. Assume that \(-2\log \ell = 463.1\) for this model, and consult the \(\chi^2\) table attached.

\[ \text{full model: } -2\ell = 402.1\]
\[ \text{reduced: } 463.1\]

\[ LR: 463.1 - 402.1 = 61 \sim \chi^2 \text{ where the reduced model is true}\]

The 95th percentile of a \(\chi^2\) is 9.49. Thus, reject the hypothesis that the reduced model is as good as the full. (p-value = 0)
Problem 3: An entomologist is studying the hatching of fruit fly eggs under a variety of conditions. Of interest to him are the temperature at which the eggs are held, and the amount of light available. He has selected three temperatures (L, M, and H) and is comparing complete darkness with ambient light. To perform the experiment, he has six sets of parents, which have laid their eggs. He takes the eggs from each parent, places them into six vials (100 eggs per vial), and places the vials into the six different sets of conditions. After two weeks the number of hatched flies in the vial is counted. He has designed the experiment in this manner as he believes that there may be a large difference in the viability of eggs from different parents, but he believes that the conditions will affect all eggs the same.

(?) pts.) What type of model do you believe would be appropriate for this data? Be sure to include the link function which you believe to be appropriate.

\[
\log_2 (\mu_{ij}) = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijk}
\]

\(i =\) temperature (3) \(j =\) light (2) \(k =\) parent (16)

This is a generalized linear mixed model (a generalized logistic regression) with \(\epsilon \sim N(0, \sigma^2 \mathbb{I})\) with observations on the same parent being correlated.

(?) pts.) Describe the linear portion of the model. Be sure to indicate which factors are fixed and which are random.

See the logistic regression model above.

\(\mu, \alpha_i, \beta_j, (\alpha \beta)_{ij}\) are all fixed. The correlated errors are random.

(?) pts.) Look at the attached output. Interpret the fixed effects values for this model. In particular, talk about which sets of conditions appear to cause the largest fraction of the eggs to hatch.

Based upon the GEE results, it appears that temperature is significant, but both the amount of light and the interaction are not.

Larger values of the parameter estimates correspond to larger odds ratios and larger probabilities. Thus, it appears that eggs hatch best at the "M" temperature.
Problem 4: Consider data from one of the following two distributions:
\[ p(y) = \begin{cases} .25, & y = 0 \\ .5, & y = 1 \\ .25, & y = 2 \end{cases} \quad \text{or} \quad p(y) = \begin{cases} .10, & y = 0 \\ .80, & y = 1 \\ .10, & y = 2 \end{cases} \]

(?) pts.) Find the Bayes discriminant rule for allocating an observation to one of the two populations.

Bayes rule says allocate an observation to the population it is most likely to come from

\[ y = 1 \rightarrow \text{Allocate to population 2} \]
\[ y = 0, y = 2 \rightarrow \text{Allocate to population 1} \]

(?) pts.) Find the discriminant rule when the probability of being in the second population is four times that of the first population.

\[ \Pi_1 + \Pi_2 = 1 \]
\[ \Pi_2 = 4 \Pi_1 \rightarrow \Pi_1 = .2, \quad \Pi_2 = .8 \]
\[ y = 1 : \quad \Pi_1 f_1 = .1, \quad \Pi_2 f_2 = .4 \rightarrow \text{Allocate to pop. 2} \]
\[ y = 0 : \quad \Pi_1 f_1 = .05, \quad \Pi_2 f_2 = .08 \rightarrow \text{pop. 2} \]
\[ y = 2 : \quad \text{same} \]

\[ \rightarrow \text{rule says assign all observations to population 2} \]

(?) pts.) What are the problems with the rule above?

This rule is “silly” as we know that 20% of all observations are from population 1, but we assign all observations to population 2.

(Despite this problem, there is no rule which does better \( \pi \)).
Problem 5: Consider a model of the form
\[ f(y|\alpha, \beta) = \frac{1}{\sqrt{\pi\beta}} \exp\left( y\alpha - \alpha^2\beta - \frac{y^2}{2\beta} \right) \].

(5 pts.) Write the distribution above in exponential form. Be sure to define \( a(), b(), c() \).

\[ f(y|\alpha, \beta) = \exp\left[ y\alpha - \frac{\alpha^2\beta}{4} - \frac{y^2}{2\beta} - \ln\left( \frac{\beta}{\sqrt{\pi}} \right) \right] \]

\( \Theta = \alpha \quad a(\phi) = 1 \quad b(\Theta) = \frac{\alpha^2\beta}{4} = \frac{\Theta^2\beta}{4} \quad c(y, \phi) = \frac{y^2}{\beta} - \ln \frac{\beta}{\sqrt{\pi}} \)

(5 pts.) What is the mean of this distribution? How about the variance?

\[ E(y) = b'(\Theta) = \frac{\Theta^2\beta}{2} = \frac{\alpha^2\beta}{2} \]

\[ \text{Var}(y) = a(\phi) b''(\Theta) = 1 \cdot \frac{\beta}{2} = \beta/2 \]

BONUS: (5 pts.) This is a reparameterization of a common model. What is that model, and how do \( \alpha \) and \( \beta \) relate to the parameters of the other distribution.

Notice that this looks rather like a normal distribution, and it has the property that \( \Theta = \alpha \).

Let \( \mu = \alpha \beta/2 \) and \( \sigma^2 = \beta/2 \). \( \Rightarrow \beta = 2\sigma^2 \)

and \( \mu = \alpha \sigma^2 \) \( \Rightarrow \alpha = \mu/\sigma^2 \).

Plugging in these definitions above,

\[ f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^4} - \frac{y^2}{2\sigma^2} \right) \]

\[ = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{y^2}{2\sigma^2} + \frac{2y\mu}{2\sigma^2} - \frac{\mu^2}{2\sigma^4} \right) \]