

MTH 325, Exam #2, Spring 2023 Name _____

Instructions: Answer each question thoroughly. For questions in Part 1, use the work you did at home to answer the questions. Be sure to answer each part of each question. In Part 2, report exact answers unless directed to round.

Part I:

Use the work you did at home to answer these questions about boilers in our dataset.

1. Write the model from your two-way ANOVA or glm model including the interaction term. Be sure to explain which level is considered the default in your two binary variables.
2. Briefly describe any boxplots, residual plots or normal plots you created to verify your model.
3. Write the equation of your logistic model below. You can write it in the form $\ln\left(\frac{p}{1-p}\right) =$ *linear model*.
4. Interpret the slope (of Worker Hours) in the context of the problem.
5. Explain the meaning of the null and residual deviance for your model in this context.

6. What is the accuracy of your model?

7. Does your confusion matrix suggest any potential problems with the data? Could masking or bias be a potential issue?

Use the data on electric bills to answer the following questions.

8. Describe the type of non-linear (parametric) model that would seem appropriate for this data. Why? Write the equation of your model.

9. What is the R^2 value for your model?

10. What is the residual standard error of your model?

11. Test your model assumptions using your residual plots and other diagnostic plots. Do they appear to be approximately satisfied? Identify any potential outliers.

12. (Bonus) Describe the LOESS model and compare it to your polynomial model. Describe any advantages or disadvantages to this model.

Use the employee data to answer the following questions.

13. Write the equation of your LASSO model below.

14. Write the equation of the linear model with the same variables below.

15. Compare the coefficients in your two models. How do they differ?

16. Are any of the retained variables in your model unable to pass a hypothesis test for the coefficient in the linear model? Explain how you would handle this in an analysis.

17. Even though the departments were encoded as ordinal variables, why could we not analyze them in the model this way?

Part II:

18. Describe at least two reasons why someone might want to create a 2^p factorial design experiment.

19. Describe one reason why we might want to recode a continuous variable as discrete?

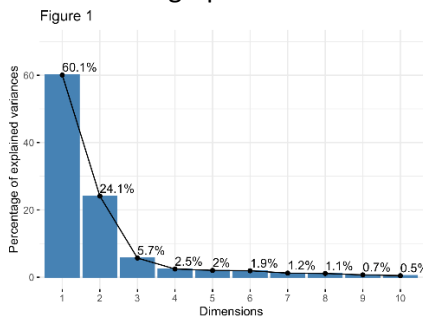
20. Describe how k-fold cross validation works in validating a model.

21. Describe unsupervised learning (in machine learning), and give an example of a machine learning algorithm that implements this learning method.

22. Describe how spline regression works in general terms.

23. How does adding a penalty improve model selection in regression? What is a potential disadvantage?

24. An example of a scree plot is shown below. How many factors should be selected for the model based on this graph?



25. Describe one advantage and one disadvantage of ensemble methods in machine learning.

26. Gaussian process regression is especially useful for uncertainty quantification. What is one disadvantage of this regression method?