MTH 325, Final Exam, 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 motels.

1. Based on your correlation table or correlation plot, identify the variable that has the highest negative correlation with Operating Margin. What is the (approximate) correlation value?

2. What is the best variable we can use to predict whether or not the motel has an indoor pool?

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. Write your confusion matrix here, and state the accuracy of your model.

5. What is the (approximate) area under your ROC curve?

6. Write the equation you obtained from your backward selection process for predicting operating expenses. Be sure to clearly indicate what each variable in the equation represents.

7. Describe how your other model selection methods differed (or were similar to) the results obtained from the backward selection process.

8. What percentage of the variability in Operating Margin can be explained by the relationship to the other model variables?

9. Answer this question and the remaining questions in Part 1 using the backward selection model you found by hand. Do your diagnostic plots suggest any outliers or model problems? Explain.

10. How do your predictions for the 5 possible locations differ? Should they have an indoor pool or not? What is the best location based on the predicted operating margins? (Give the location number.)

11. Interpret the meaning of the Indoor Pool coefficient in the context of the problem.

12. If you could only model operating margin with a single variable, what would it be, and why?

Use the work you did at home to answer these questions about the time series model.

13. Does the model appear approximately stationary or does there appear to be a trend? Consider any boxplots or histograms here, as well as any time series plots or decompositions you may have done.

14. Based on your ACF graph, how many lags should be included in your time series model? Why?

15. What settings did you use for your ARIMA model? Why? What diagnostics did you use to select these settings?

16. Write the equation of your final time series model.

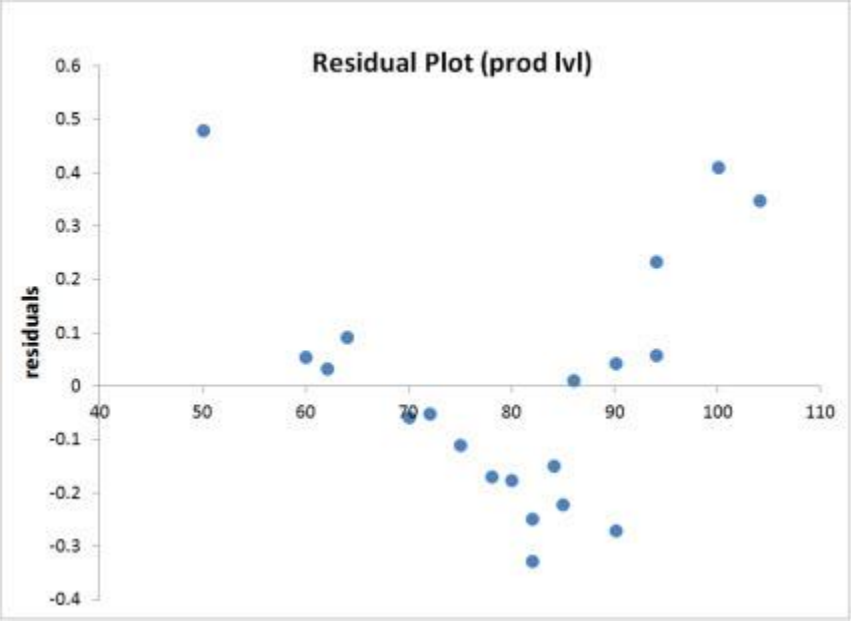
17. What is the AIC of your final model? How good does the model appear to fit?

Part II:

18. Recall that $Cov(X,Y) = E(XY) - E(X)E(Y)$. For the probability density function $f(x,y) = \frac{25}{192}x^{3/2}y^{2/3}, y \in [0,1], x \in [0,4]$, find the covariance.

19. Consider the small data set $\{(2,1),(5,3),(8,7)\}$. Find the value of the regression coefficients for $y = \beta_0 + \beta_1 x$, using the normal equation $(A^T A)^{-1} A^T Y = B$. Write the coefficients you find in the equation.

20. Examine the residual plot below. Identify some potential issues with the linear model used to produce these residuals. (For example, without knowing the variables names, use "independent variable" for x, and the "dependent variable" for y.)

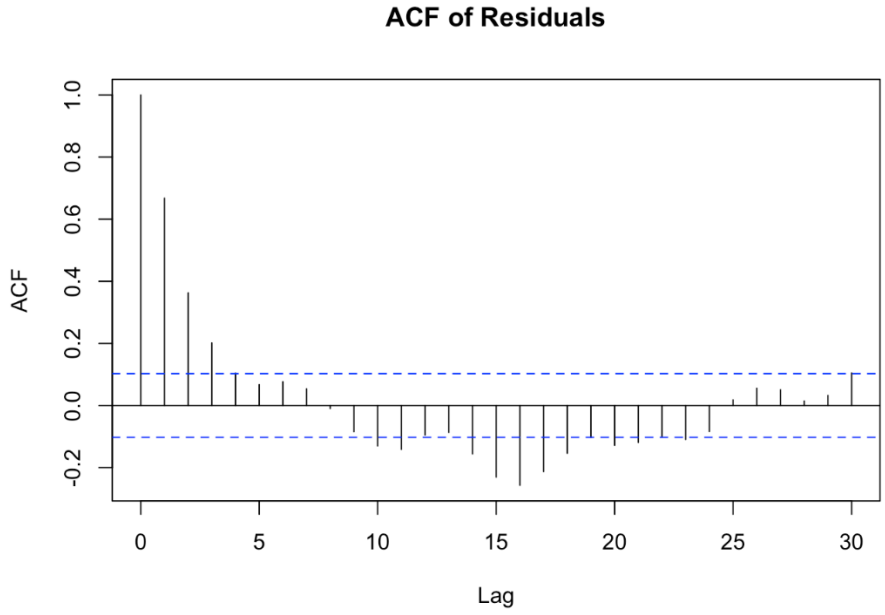


21. Describe clustering (in machine learning), and give an example of a machine learning algorithm that implements this learning method.

22. Describe how LOESS regression works in general terms.

23. What are some reasons it might be beneficial to use a non-parametric nonlinear model for a regression problem rather than a parametric non-linear model?

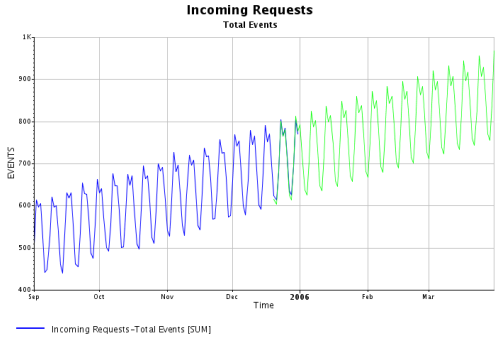
24. An example of an ACF plot is shown below. How many lags should be used in an ARIMA model based on this graph?



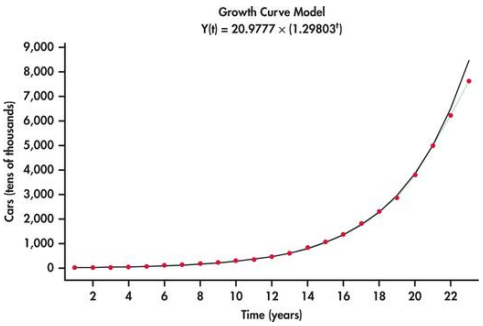
25. What are some properties of seasonal time series?

26. For each of the graphs of time series below, identify the type of trend. Options include: stationary (no) trend, exponential trend, linear trend, polynomial trend, random walk, log trend.

a.

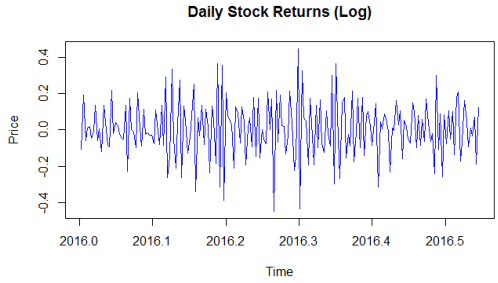


b.



c.

d.

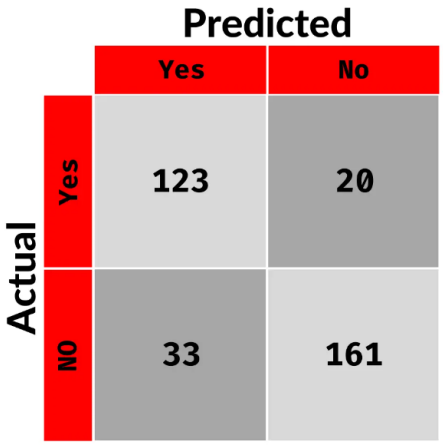


27. Explain why autocorrelation prevents us from using traditional regression to model some time series data.

28. Why are irregular time series so much more difficult to work with than regular time series? Describe some methods we can use to make irregular time series more regular.

29. Describe each of the components of a SARIMA model.

30. A confusion matrix is shown below. What is the accuracy of this model?



31. How do we use the AUC (of an ROC curve) as a diagnostic for a classification model?