ISYE 6414
Spring 2021

PRACTICE FINAL EXAM PART 1 SOLUTIONS

PART 1: TRUE/FALSE

1. **False** [3.11. Assumptions and Diagnostics] If constant variance or normality assumptions do not hold, we apply a Box-Cox transformation to the response variable.
2. **False** [3.13. Model Evaluation and Multicollinearity] Multicollinearity means there is a dependency between predicting variables which would equate to the columns in the design matrix.
3. **False** [4.5. Statistical Inference]
   - When R-squared is used as explained variability: The denominator of the ratio can be thought of as the total variability in the dependent variable, or how much y varies from its mean. The numerator of the ratio can be thought of as the variability in the dependent variable that is predicted by the model. Thus, this ratio is the proportion of the total variability explained by the model.
   - In logistic regression, the response variable is binary. So the classic R-squared does not make sense in measuring explained variation.
4. **False** [Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation] + [Topic 4.2 Lesson 5: Logistic Regression: Statistical Inference]
   - Interpretation of the regression coefficients of Poisson regression is in terms of log ratio of the rate.
   - Interpretation of the regression coefficients of Logistic regression is in terms of log odds.
5. **True** [Topic 4.5 Lesson 17: Poisson Regression: Statistical Inference] In Poisson regression, if the sample size small, the statistical inference is not reliable. Thus, the hypothesis testing procedure will have a probability of type I error larger than the significance level.
6. **False** [5.7. Regularized Regression: Approaches] Because LASSO can eliminate predicting variables using the penalty while Ridge and Elastic Net retain coefficients, LASSO will have the same number or LESS predicting variables.
7. **False** [5.12. Variable Selection] R-squared is not compared during stepwise variable selection. Variables are selected if they reduce the AIC or BIC of a model.
8. **True** [5.3. Prediction Risk Estimate] Adding more variables will increase the variability and possibly induce multicollinearity. Adding more variables also reduces the bias in the model

since it has an additional predictor to conform to which keeps the model from favoring one of the original predictors.

9. **False** [5.4. Model Search] We desire the model that has the smallest AIC or BIC.
10. **True** [5.3. Prediction Risk Estimate] BIC penalizes complexity more than other approaches.
11. **False** [5.3. Prediction Risk Estimate] For linear regression under normality, the variance used in the Mallow's Cp penalty is the estimated variance from the full model.
12. **True** [5.4. Model Search] ”Stepwise regression is a greedy search algorithm. It does not guarantee to find the model with the best score.”
13. **True** [5.7. Regularized Regression: Approaches] Ridge regression has been developed to correct for the impact of multicollinearity. If there is multicollinearity in the model, all predicting variables are considered to be included in the model but ridge regression will allow for re-weighting the regression coefficients in a way that those corresponding to correlated predictor variables share their explanatory power and thus minimizing the impact of multicollinearity on the estimation and statistical inference of the regression coefficients.
14. **False** [5.7. Regularized Regression: Approaches] Elastic net often outperforms the lasso in terms of prediction accuracy. The difference between lasso and elastic net is the addition of a penalty just like the one used in ridge regression. By considering both penalties, L1 and L2 together, we have the advantages of both lasso and ridge regression.
15. **False** [4.9. Classification] Random sampling is computationally more expensive than the K-fold cross validation, with no clear advantage in terms of the accuracy of the estimation classification error rate. K fold cross validation is preferred at least from a computation standpoint.

PART 2: MULTIPLE CHOICE

16. **B**: -0.948**;** t-stat = (Coefficient-comparison value)/SE = (41.33160-50)/9.13907 = -0.948
17. **B**: One; [3.3.11: Assumptions and Diagnostics] There are 27 observations hence the rule of thumb is 4/n = 4/27 = 0.15. From the graph there is only one observation with a Cook's distance measurement >0.15.
18. **C**: Increase of 20.284; [Topic 3.2 Lesson 6: Inference for Regression Parameters] The estimated coefficient for doserate is 20.284.
19. **C**: 3, 23;
    - F(p, n-p-1) = F(3, 23)
    - n = # of observations=27
    - p = # of predicting variables=3
    - p = 3
    - n-p-1 = 27-3-1=23

20. **C**: 0.6697; [Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation]

    - $$E(Y|X_{math} = 45.5, X_{langarts} = 50, X_{male} = 0) = \exp \\ (2.68766 - 0.003523 * (45.5) - 0.012152 * (50) - 0.400921 * (0)) = 6.819386 \\ E(Y|X_{math} = 45.5, X_{langarts} = 50, X_{male} = 1) = \exp \\ (2.68766 - 0.003523 * (45.5) - 0.012152 * (50) - 0.400921 * (1)) = 4.566963 \\ E(Y|X_{math} = 45.5, X_{langarts} = 50, X_{male} = 0) - E(Y|X_{math} = 45.5, X_{langarts} = 50, X_{male} = 1) \\ = 2.252423$$

21. **D**: Decrease by 1.21%; [Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation] The estimated coefficient for langarts is -0.012152. A one unit increase in langarts gives us exp(−0.012152) = 0.9879215. It is interpreted as the expected number of days missed decreasing by 1.21% (1-0.9879). Hence, given that the other predictors in the model are held fixed, one unit increase in langarts results in the expected number of days missed decreasing by 1.21%.
22. **A**: 6.8773; [Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation]

    - $$E(Y|X_{math} = 50, X_{langarts} = 48, X_{male} = 0) = \exp \\ (2.68766 - 0.003523 * (50) - 0.012152 * (48) - 0.400921 * (0)) = 6.877258$$

23. **D**: 3; [Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation] The coefficients of intercept, langarts and male are statistically significant at α=0.05.
24. **B**: Chi-Squared, 312; [Topic 4.5 Lesson 20: Poisson Regression: Goodness of Fit Assessment Data Example] Hypothesis Testing Procedure: Under null hypothesis, the approximated distribution of the deviance test statistic D~χ^2 with df = n-p-1 = 316-3-1 = 312, where n = # of observations, p = # of predicting variables.
25. **C**: L1, LASSO, forcing beta coefficients to zero; [5.7. Regularized Regression: Approaches] Ridge regression does NOT perform variable selection because the variables remain in the model. LASSO performs variable selection because coefficients will be forced to 0, which removes the variable from the model. LASSO represents the L1 penalty.
26. **D**: 32; A model with p predictors has 2^p different model combinations.