Practice Exam Part A

Short Answer Questions

1. Why is evidence informed reasoning so important for systems engineering?
2. What steps do we take to implement evidence informed reasoning?
3. List two sources of evidence
4. What is the difference between observations and experiments?
5. Define a well-formed hypothesis.
6. What is the difference between cross sectional data and time series data?
7. What is the most common problem in data visualization?
8. Compare and contrast a histogram and a density plot.
9. What is the center line or point in the box on a box plot?
10. What are the observations above or below the whiskers in a box plot?
11. What is a Q-Q plot and why is it useful?
12. In R what plotting technique is particularly useful for visualizing a categorical variable vs. a quantitative variables?
13. What is the typical graphical display in two dimensions?
14. What display shows pairwise relations in more than two dimensions
15. What is the center line or point in the box on a box plot?
16. The first principal component is a projection that maximizes what attribute of the data?
17. How many principal components are there in a data set with only quantitative variables?
18. What is the linear algebra solution for the first principal component?

19. Why do we often use a correlation matrix and not the covariance matrix to compute the principal components?

20. In the simple linear regression model $Y = \beta_0 + \beta_1 X_1 + \epsilon$, name all of the terms in the model. (Name Y, β_0, β_1, X_1, and ϵ?)

21. In the linear regression model $Y = \beta_0 + \beta_1 X_1 + +\beta_2 X_2 + \epsilon$, how many coefficients and parameters are there?

22. What are the assumptions for the error term in a multiple regression?

23. What is the sum of squares decomposition of the total sum of squares used in regression analysis?

24. What does $\hat{y}_i$ mean in regression analysis?

25. What is the hypothesis for a t-test in multiple linear regression?

26. What is the hypothesis for the F-test (model utility test) in multiple linear regression?

27. What is the hypothesis for the partial F-test in multiple linear regression?

28. Can we use graphical methods for model diagnostics?

29. What is the most important graphic used in diagnosing or evaluating regression results?

30. What diagnostic plot do we use to check the Gaussian assumption for the errors?

31. Write a main effects model for a regression with two predictor or explanatory variables, X_1 and X_2. X_1 is quantitative but X_2 is qualitative with three levels: $L_1, L_2,$ and L_3. We want L_1 to be the base case.

32. What plot do we use to determine if an observation is influential?

33. What plot helps us detect heteroscedasticity or non-constant variance?

34. When do we consider transformations of the response variable?
35. What is the goal of the Box-Cox method?
36. What does a pairwise interaction term mean?
37. Give examples of nonlinear transformations that are permitted for the predictor variables in multiple regression?
38. Give examples of nonlinear transformations that are permitted for the coefficients of the predictor variables in multiple regression?
39. Analysis of Covariance (ANCOVA) combines what types of predictor variables?
40. For a qualitative variable with four levels, red, blue, orange, and yellow, give a dummy or treatment encoding with orange as the base case.
41. Write a main effects model for a regression with two explanatory variables, X_1 and X_2.
42. Write a main effects plus interaction effects model for a regression with two explanatory variables, X_1 and X_2.
43. What test do we use for the hypothesis that a qualitative variable with 5 dummy variables is correlated with the response in a regression with four other variables?
44. Name one metric for prediction evaluation.
45. What is the method for evaluating predictive performance that allows us to generalize to new data.
46. What is the method for evaluating predictive performance when we don't have enough data.
47. List methods to use for model selection and validation in multiple regression.
48. Does stepwise regression give the optimal solution for model selection?
49. What does a large value for leverage mean?

50. Why don't we use t-tests for variable selection?

51. If p is the probability of an event, what are the odds of the event in terms of p?

52. Instead of modeling the event (i.e., 0 or 1) logistic regression uses something else for the response variable. What is it?

53. Why don't we use multiple linear regression for a qualitative response?

54. The generalized linear modeling technique called logistic regression is used for problems with a binary response. However, instead of modeling the binary response, what does logistic regression model and provide as its output?

55. Is a logistic function linear or nonlinear?

56. Is the logit?

57. Logistic regression makes a transformation of the response to form a problem with a linear or nonlinear expression of the regression parameters on the right-hand side?

58. Logistic regression replaces sum of square errors used in multiple regression with what?

59. What restrictions do we have on the scales for the predictor variables in logistic regression?

60. What is the hypothesis for the χ^2, likelihood, or deviance test for a logistic regression model?

61. What is the null model in logistic regression?

62. $exp(\beta_i)$ gives what meaning for logistic regression.

63. Define true positive.

64. Define true negative.

65. What are the axes of an ROC curve.