Data Science Question list

1. State the data science work suggested by Jeff Hammerbacher’s Model.
2. List the 9 Top Programming Languages for Data Science reported on May 24, 2021, by the edX team.
3. List the basic three principles of DOX. Explain how and why to use the principles.
4. Construct a two-factor fixed effect ANOVA table. Explain how to test the factor effects including the main effects and the interaction effect.
5. Explain the typical content of a Measurement Systems Capability Study. How to calculate the gauge capability?
6. List 4 types of probability sampling methods. Give a brief explanation of each one.
7. List 4 types of non-probability sampling methods. Give a brief explanation of each one.
8. List 8 types of plots for data visualization. Give a brief explanation of each one.
9. List two methods to deal with overplotting in scatter plots.
10. List 5 criteria for model selection. Give a definition or description of each criterion.
11. Draw a general picture of the bootstrap procedure. What are the typical situations the bootstrap will be used? Can the bootstrap estimate prediction error? Why?
12. List the three classes of model selection methods. Give a description of each class.
13. List the formula for LASSO. When the LASSO will be helpful?
14. List two linear methods for classification problems. Give a definition or description of each method.
15. List two non-linear methods for classification problems. Give a definition or description of each method.
16. Give the definition of a GAM for classification.
17. Give the tree algorithm.
18. Give the boosting algorithm for the regression tree.
19. Give the formula for a support vector classifier that maximizes a soft margin.
20. Give two methods to deal with a multiple testing problem by better controlling the FWER and FDR, respectively.