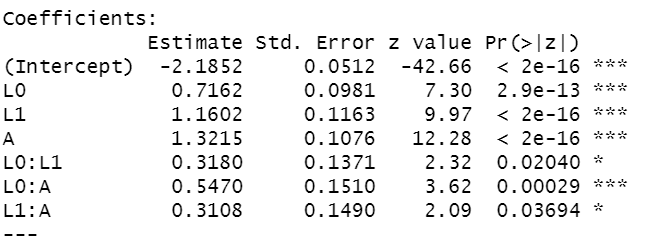
Biostatistics Assignment

1. Given the below logistic regression model output, what would be the predicted probability that the outcome is 1 given covariate values L0=0, L1=1, A=1?

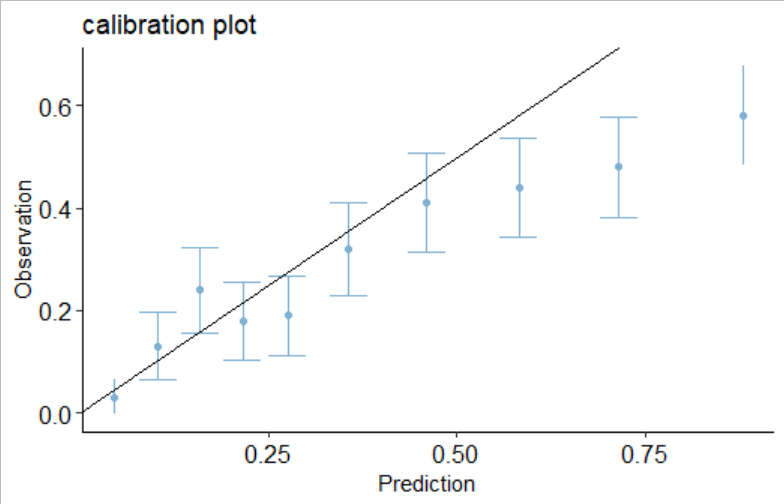
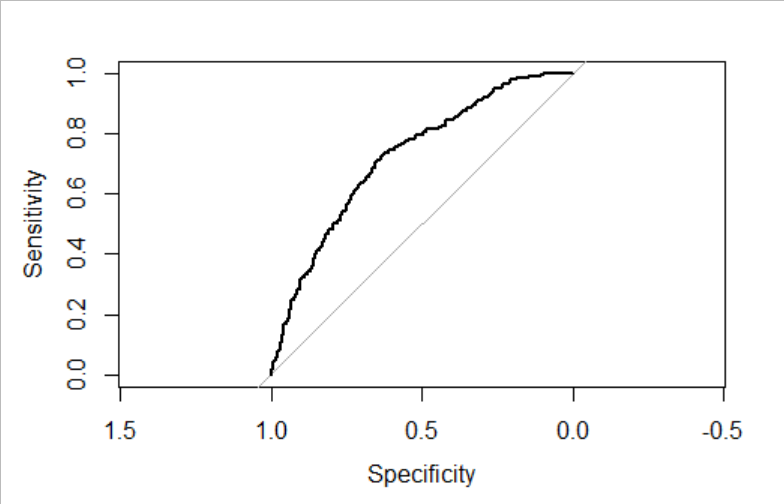


1. Suppose that exchangeability and positivity hold conditional on L0 and L1 with A as the treatment and Y as the outcome. In the data used to fit the model from question 1, Pr(L0=0,L1=0)=.45, Pr(L0=0,L1=1)=.05, Pr(L0=1,L1=0)=.1, and Pr(L0=1,L1=1)=.4. What is the estimated average treatment effect (E[Y(1)]-E[Y(0)]) using standardization and the model in 1? Show your work.
2. Below is a table of the frequency with which each set of values appeared in a dataset with covariates L0 and L1, treatment A, and outcome Y. Under exchangeability and positivity, what is the estimated effect (E[Y(1)]-E[Y(0)]) using IPW? No need for logistic regression to get probabilities, you can just use proportions from the table. Show your work.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| L0 | L1 | A | Y | N |
| 0 | 0 | 0 | 0 | 3666 |
| 0 | 0 | 0 | 1 | 405 |
| 0 | 0 | 1 | 0 | 303 |
| 0 | 0 | 1 | 1 | 137 |
| 0 | 1 | 0 | 0 | 268 |
| 0 | 1 | 0 | 1 | 105 |
| 0 | 1 | 1 | 0 | 64 |
| 0 | 1 | 1 | 1 | 99 |
| 1 | 0 | 0 | 0 | 643 |
| 1 | 0 | 0 | 1 | 156 |
| 1 | 0 | 1 | 0 | 88 |
| 1 | 0 | 1 | 1 | 115 |
| 1 | 1 | 0 | 0 | 1205 |
| 1 | 1 | 0 | 1 | 1203 |
| 1 | 1 | 1 | 0 | 149 |
| 1 | 1 | 1 | 1 | 1394 |

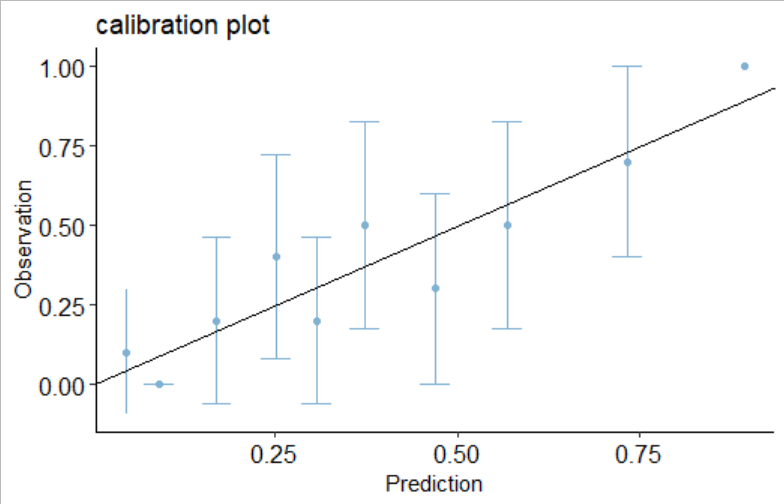
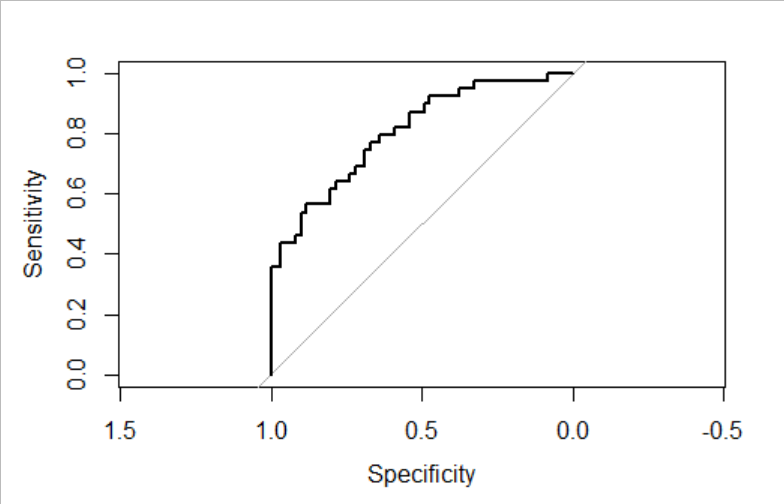
1. Below are two sets of calibration and ROC plots for the same logistic regression model. One set was generated from the data to which the model was fit. Another set was evaluated on a holdout test dataset that was not used to fit the model.
2. Set 1 comes from the data used to train the model
3. Set 2 comes from the data used to train the model
4. There is not enough information to tell

Explain your answer.



AUC=.72

**Set 1**



AUC=.81

**Set 2**

1. Suppose disease prevalence and age breakdown in a sample is given by the table below

|  |  |  |
| --- | --- | --- |
| Age bracket | Sample Proportion in Age Bracket | Sample Disease Prevalence in Age Bracket |
| 65+ | .4 | .15 |
| <65 | .6 | .09 |

What is the overall disease prevalence in the sample? Suppose in the population of interest the age breakdown is 30% 65+ and 70% <65. What is the estimated disease prevalence in the population of interest from this sample?

1. The outcome Y is missing ½ of the time when L=0 and ¼ of the time when L=1. The mean value of Y when it is observed and L=0 is 5. The mean value of Y when it is observed and L=1 is 10. Pr(L=1)=1/2.
2. What is the mean *observed* value of Y?
3. What is the *overall* mean value of Y including observed and unobserved values? Show your work.