1. Default
   1. Done
   2. 2.8% test error
   3. Error rates
      1. 2.9%
      2. 2.7%
      3. 3.9%
   4. There seems to be a slight improvement in the test error rate when the dummy variable for student is included in the logistic regression overall.
2. Bootstrapping on Default
   1. Std. Errors
      1. Intercept=0.4348
      2. Balance =4.985\*10^-6
      3. Income=2.274\*10^-4
   2. Done
   3. Done
   4. The standard error for the intercept is roughly the same between the two. The standard error for balance is lower, but the standard error for income is higher.
3. LOOCV
   1. Done
   2. Done
   3. This observation was correctly predicted.
   4. Done
   5. The LOOCV estimate for the test error is about 48.85%. This indicates that the logistic regression model involving only lag1 and lag2 as predictors is only a little bit better than random guessing (50% error rate), indicating that there is little to no correlation in the data between these 2 predictors and the direction of the market.
4. Cross-Validation Testing
   1. In this data set, *n*=100, *p*=1. The equation of this model is y=B0+B1x+B2x^2+e, where B0=0, B1=1, B2=-2, and e is the irreducible error.
   2. The plot of y against x closely resembles the parabola of y=-2x^2+x, with the curve of best fit likely being very close to said parabola.
   3. LOOCV errors
      1. Poly1=7.288
      2. Poly2=0.937
      3. Poly3=0.957
      4. Poly2=0.954
   4. Regardless of what random seed is used, the errors are the same. This is because when LOOCV, every possible group of n-1 training observations is used exactly once as a training set, leaving neither randomness in the groups nor randomness in the choice of which groups to use, as they are all used. Therefore, there is no element of chance or random error in LOOCV, meaning that setting any seed will generate the same results.
   5. The LOOCV error was the smallest when the second degree polynomial was used. This makes sense, as the true relationship between x and y is a second degree polynomial, and neither a cubic nor a quartic would be able to model it as well as another of the same degree.
   6. The statistical significance of the coefficient estimates, based on the p-values, matches with the conclusions drawn based on the cross-validation results. In the first degree polynomial, the linear term had a high p-value, indicating it to be rather insignificant, as it was of the wrong degree and could not model a non-linear relationship effectively. The second degree polynomial had a low p-value for the first degree term and a near-0 p-value for the second degree term. This makes sense as well, as there is a linear and quadratic term in the true relationship. The cubic and quartic terms have extremely high values, however, which fit since they cannot be used to accurately predict a quadratic model.
5. Boston
   1. Mean of medv=22.533
   2. Standard error of medv=0.4089
   3. Bootstrap standard error=0.4107. The bootstrap standard error estimate is higher than the original estimate, but is likely in fact more accurate due to the original estimate using only a single sample of the population, while the bootstrap simulates many different samples from the same sample, in order to reduce variance.
   4. The bootstrap confidence interval is about the same width as the one from t.test, but it is shifted slightly lower in both boundary values.
   5. Median=21.2
   6. The bootstrap standard error of the median is 0.3778. It is noticeably lower than any of the estimated mean errors, due to how the median is not affected nearly as much by all of the values in the data set, making it less variable than the mean.
   7. Estimated tenth percentile=12.75.
   8. Bootstrap standard error of the tenth percentile is 0.4768. It appears to be higher than all of the other estimated standard errors due to the high volatility of the 10th percentile, especially if there is a large amount of observations in the data set. Changing even a single value in that case would affect every percentile.