# Case Study 2

Sam Lee & Patric Platts

## 1

After standardizing the numerical covariates, we will fit two variable selection models: We will fit a lasso penalized regression model and a partial component regression (PCR) model. After fitting our models, we will apply a level of k-fold cross-validation to select the best hyperparameters for the model of best fit. After selecting our chosen model, we would examine the covariates that are the largest in absolute value (since the model would fit on scaled vectors) and which are significant.

## 2

We would look at the adjusted $R^2$ statistic as an overall model fit adjusted for the number of covariates included. We would also compare how well our selected covariates from the lasso regression compare to the full model fit by running an ANOVA test of the two models (comparing their f-statistics of significance).

## 3

To assess the predictive power of the model, we would look at the (out-of-sample) RMSE through k-fold cross-validation. To specifically look at the predictive power of the covariates we included, for a model like PCR, we could assess the portion of variance explained by each component as well as the total cumulative explained variance[1]. For our lasso regression model, in addition to RMSE as assessed through cross-validation and the adjusted $R^2$, we would create ($B$ # of) bootstraps of the data and fit the lasso model on each bootstrap, checking the frequency at which each covariate is selected[2]. We can also use bootstrap to obtain confidence intervals for each coefficient as see if these are statisticaly significant.

[1] We would assess this using the criteria established by the following expression: $\sum_{j=1}^{k} \frac{\lambda_j^2}{\sum_{i=1}^{p} \lambda_i^2}$, where each $\lambda_j$ corresponds to the singular value of the $j$-th principal component.

[2] We can assess the selection frequency of each $j$-th covariate by using this bootrstrapping method as described through the following expression: $\frac{1}{B} \sum_{b=1}^{B} \mathbb{1}(\hat{\beta}_j^{\text{lasso, } b} \neq 0)$.