**Problem 1:**

1. Part 2 and Part 3 are not equivalent because they make different assumptions about the relationships between predictors and the response variable for males and females. Part 2 assumes that the relationships are the same for both genders, while Part 3 allows for different relationships by fitting separate models.
2. See r script for code.

Using interaction terms lm(charges~age+bmi+gender+age\*gender+bmi\*gender,data=insurance)

males: Yhat = -8012.79 + 238.63\*age + 409.87\*bmi

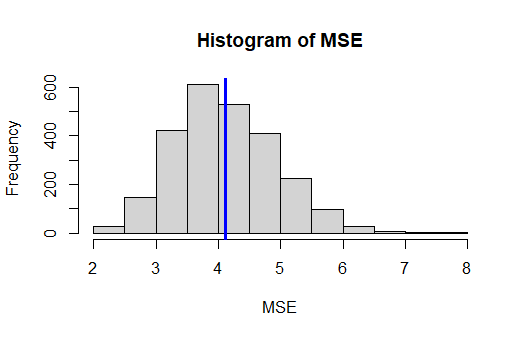
females: Yhat = -4515.22 + 246.92\*age + 241.32\*bmi

This is the same as part 3.

1. A model with interaction terms returns a better r2 value than without interaction terms. For no interaction terms, r2 was 0.120258. For interaction terms, r2 was 0.122032. The higher r2 is better since it represents smaller differences between the observed data and the fitted values.

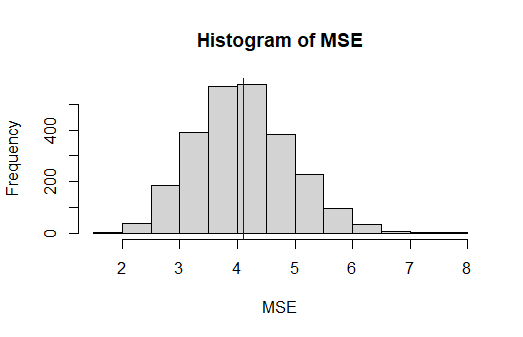
**Problem 2:**

1. Yes, multicollinearity is a problem for making accurate predictions. If there is multicollinearity, the standard errors for the least squares estimates could be very large. These errors could skew the predictions produced by the model and lead to less accurate predictions
2. Correlation between x1 and x2 is 0.9404249
3. Test MSE = 4.173413
4. Mean = 4.108063



The histogram is relatively bell shaped around the average test MSE. There are some high outliers.

1. Correlation between x1 and x2 is 0.03316596
2. Mean = 4.099373



The histogram is very bell shaped around the mean test MSE. This histogram and the one from part d look pretty similar with some high outliers.

1. Multicollinearity is a problem based on our simulation studies. The different models of multicollinearity vs no multicollinearity gives us different MSEs and average MSEs after multiple iterations.

**Problem 3:**

1. Data set split into a training and test set in R script.
2. Ridge regression applies a penalty to the coefficients. So scale is important for regularized models. Without scaling, the penalty could have different effects on each coefficient and the coefficient would vary in size.
3. Optimal λ = 0.01
4. Value of l2 = 534.899
5. test MSE: 1416737
6. Optimal λ = 0.01
7. Value of l­1 = 624.8116
8. test MSE: 1417067
9. Based on the results, both ridge and lasso regression models provide reasonably good predictive performance for the number of college applications received. Ridge regression has a slightly lower test MSE, indicating better predictive accuracy, compared to lasso regression. But, the difference in test MSE between the two approaches is not substantial, so the choice between ridge and lasso regression would also depend on other factors like model interpretability. Ridge regression tends to perform well when many predictors are potentially relevant, while lasso can perform variable selection by shrinking some coefficients to exactly zero, making it useful when you want a more interpretable model with fewer predictors.