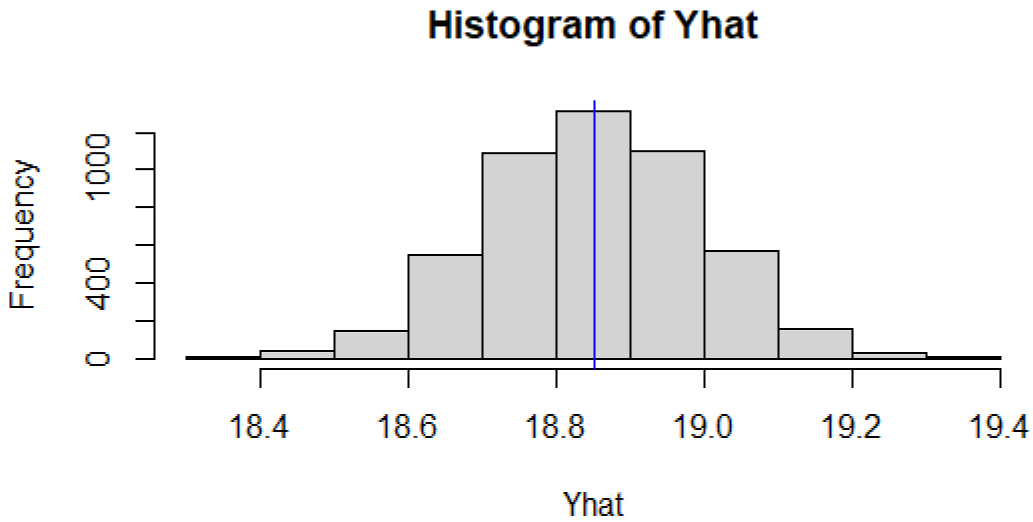
Neha Maddali

Question 1:

a. Given the same X values, prediction intervals are wider than confidence intervals because they take into account both reducible error and <u>irreducible error</u>
b. To evaluate the accuracy of the least square estimates B^ for a single dataset, we can report its <u>R-squared value (R^2)</u>
c. To estimate the regression coefficients for a polynomial regression mode, we use the technique known as <u>the least squares method.</u>
d. To diagnose whether or not multicollinearity is present in our model, we can look at the <u>variance inflation factor (VIF)</u>
e. If we carry out 150 hypothesis tests and set α = 0.08, how many type 1 errors can we expect to make? <u>8%</u>
f. If the constant variance assumption on the error term is violated, we should consider transforming <u>Y</u>
g. The presence of multicollinearity is problematic for <u>accurate predictions of the estimated coefficients (increases standard error for estimates)</u>, but it is not problematic for <u>when collinear variables are used as control variables</u>
h. Suppose the true relationship between the response and a predictor is linear. We fit a polynomial regression model (Model 1: Y~ X+X^2+X^3) and a standard linear regression model (Model 2: Y~X). We split our data into a training set and a test set and compute their training MSE and test MSE. Which model would we expect to have a smaller training MSE? <u>Model 1</u>
i. Continue with the same setup as above. Which model would we expect to have a smaller test MSE? <u>Model 1</u>
j. <u>Y^ (E(Y^) = E(Y))</u> is an unbiased estimate for E(Y)

Question 2:

a. B0 = 2, B1 = 3, B2 = 4
b. E(Y) = 12
c. E(Y) = 18.92324
d. E(Y) = 18.85062



e.
f. 18.72188
18.07899 to 19.36477 using prediction interval with 0.05 significance level

Question 3:

a. H_0: $B_1 = B_2 = B_{17} = 0$, H_1: at least one B_j is non-zero
Test statistic (F^*) = 38.27
Null distribution: $F_{17, 759}$
p-value: < 2.2e-16
Decision rule: if the p-value is less than 0.01, reject H_0
Conclusion: The null hypothesis is rejected and the results are statistically significant. B_j is significantly different from 0 at significance level 0.01.
b. H_0: $B_1 = 0$, H_1: $B_1 \neq 0$
Test statistic = 1.993
Null distribution: t-distribution with 759 degrees of freedom
p-value: 0.046605
Decision rule: if the p-value is less than 0.01, reject H_0
Conclusion: the null hypothesis is not rejected and we do not have evidence that B_1 is significantly different from 0, at a significance level 0.01.
c. A 3.3813758 regression coefficient shows there is a positive relationship between the response Grad.Rate and the dummy variable PrivateYes.
d. Var(Yi) = 5.609779
e. Multicollinearity causes standard errors for the regression coefficients to be too high, which can cause the t-statistics to be too low. With strong multicollinearity there might be a regression coefficient that is very highly significant based on the F test but for which not one of the t-tests of the individual predictors is significant. Multicollinearity should not affect the F-test