# Homework 4

Aiden Kenny
STAT GR5205: Linear Regression Models
Columbia University

November 9, 2020

## Question 1

*Collaborators:* None

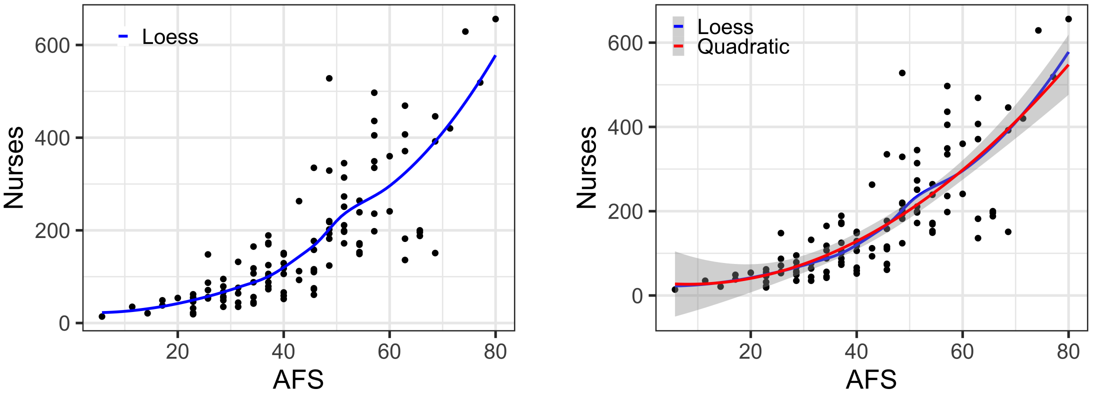

Figure 1: Overlaying a loess smoother and a quadratic polynomial to the scatterplot of AFS vs. Nurses. The left plot displays only the loess smoother, while the right plot displays both.

(a) The loess smoother is the blue line displayed in Figure 1. For clarity, the left plot displays only the loess smoother. The loess smoother indicates that a linear function does not seem plausible for this data, as the curve is too non-linear.

(b) The fitted quadratic model is given by $\hat{Y} = 33.548 - 1.666x + 0.101x^2$, and can be see in the right panel of Figure 1. As we can see, the quadratic model almost perfectly overlays the loess smoother, further indicating that a linear relationship is not plausible.

(c) Given our model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$, to determine if the quadratic term can be dropped from the model, we will test $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$. The $p$-value for this test is 0.00032, so for any reasonable significance level, we will reject $H_0$. It seems that there is a non-negligible quadratic relationship. Interestingly, if one were to re-run the same test for $\beta_0$ and $\beta_1$, the corresponding $p$-values would be 0.515 and 0.495, which would not give us grounds to reject either of those null hypotheses. That is, we could say the slope and linear term of this model are insignificant, but the quadratic term is not.

(d) When AFS is 30%, a 95% prediction interval is $\big(-89.781, 239.004\big)$, and when AFS is 60%, a 95% prediction interval is $\big(133.136, 462.503\big)$. The simultaneous confidence level for both of these intervals is $0.95^2 = 0.9025$, so we are 90.25% confident that both of these predictions will occur.

(e) The diagnostic plots have been printed in Figure 2. We can see right away that the assumption of equal variance among the residuals is violated. For both residual plots, the residuals create a cone-like pattern, where the variance in the residuals starts small, and then increases. In addition, there are several significant outliers in the QQ-plot, meaning that the assumption of normally-distributed residuals is most likely being violated as well.

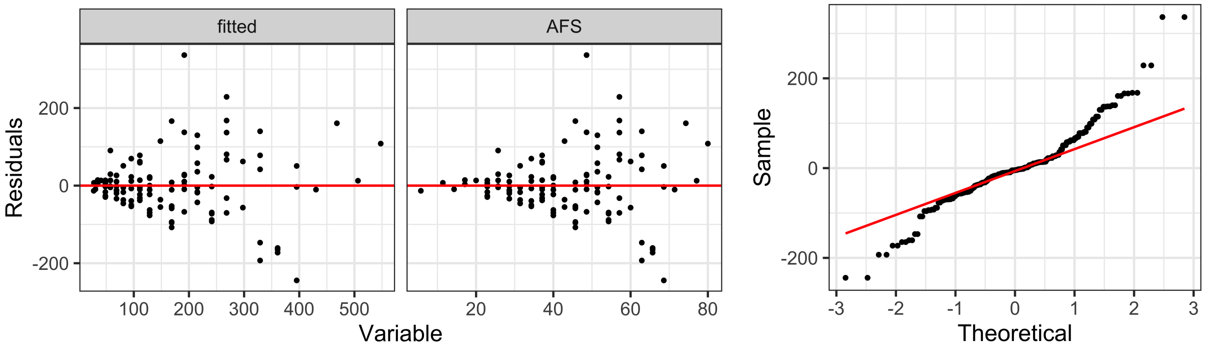

Figure 2: Diagnostic plots for the quadratic model.

## Question 2

## Question 3

## Question 4

*Collaborators:* None

(a) For this question we will assume the matrix $\mathbf{X}$ is *centered*, so that each predictor has mean zero. In $n$-dimensional Euclidean vector space, for any $\mathbf{v} \in \mathbb{R}^n$ we have $\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v}$, so the ridge regression loss function becomes

$$Q(\boldsymbol{\beta}; \lambda) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2 = \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}.$$

Differentiating the loss function with respect to $\boldsymbol{\beta}$ gives us

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + 2\lambda\boldsymbol{\beta} = -2\mathbf{X}^T\mathbf{y} + 2\big(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\big)\boldsymbol{\beta} \overset{\text{set}}{=} \mathbf{0},$$

and finally solving for $\boldsymbol{\beta}$ gives us $\hat{\boldsymbol{\beta}}_\lambda = \big(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\big)^{-1}\mathbf{X}^T\mathbf{y}$.

(b) When $\lambda = 0$ we have $\hat{\boldsymbol{\beta}}_{\lambda=0} = \big(\mathbf{X}^T\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{y}$, the OLS estimator.