Homework 14
David Yang

1. **The countries data set is from the CIA and reports Life Expectancy, Birth Rate, Real GDP per Capita and Education expenditure as a % of GDP. Find a model to predict life expectancy, or some transformation of life expectancy, from some or all of the other variables, or transformations of the other variables (interactions are possible, too). Show evidence that the Normal linear regression assumptions appear to be met. Make a 95% prediction interval for the life expectancy in a country with a birth rate of 12 births per 1000 persons in a year, GDP per capita of $\$70,000$, and with 4% of GDP spent on education (close to the values for the USA).**

   Solution. See attached R Output for specific details.

   In my linear model, I use the birth rate and GDP variables to predict life expectancy (education was left out as it was not significant, and none of the interaction variables with education were significant either). The assumptions for Normal linear regression were satisfied; there appears to be a relatively linear relationship between the variables, we see that the residuals are relatively normal, and satisfy the homoscedasticity and independence conditions.

   The 95% prediction interval I derived for the life expectancy under my model was approximately

$$[68.9, 82.6]$$

   which seems on par with the data for countries such as the UK and the US. ∎

2. **Suppose that the relation of family income to consumption is positive and roughly linear. Of those families in the 90th percentile of income, what proportion would you expect to be at or above the 90th percentile of consumption: (a) exactly 50%, (b) less than 50%, of (c) more than 50%? Justify your answer.**

   Solution. For families in the 90th percentile of income, we expect the proportionto be at or above the 90th percentile of consumption to be $\boxed{\text{(c) more than } 50\%.}$ This follows from the fact that the relation of family income to consumption is positive and roughly linear, so we expect families with higher incomes to have higher levels of consumption. ∎

3. **The *weighted* regression model assumes**

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \ \ \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{W}^{-1})$$

where $\boldsymbol{W}$ is a known, symmetric $n \times n$ positive definite matrix, and $\boldsymbol{X}$ is $n \times p$ and is also known.

(a) **Define $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Y}$. Show that**

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}).$$

Solution. We use our trick of adding and subtracting by the same constant in our left hand side expression, and then expanding. Note that

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = ((\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}))^T \boldsymbol{W} ((\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}))$$

Expanding, we find that

$$\begin{aligned}
(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) &= ((\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}))^T \boldsymbol{W} ((\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})) \\
&= (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}) \\
&\quad + (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{W} (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})
\end{aligned}$$

Note that the latter two terms are both 1×1, so they are equal. To simplify the above expression, we work with the term

$$\begin{aligned}
(\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) &= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{X}^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Y} - \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Y} - \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}((\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Y})) \\
&= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Y} - \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Y}) \\
&= 0.
\end{aligned}$$

Thus, we simplify to get that

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

as desired. ■

(b) **Explain why (a) implies that $\hat{\boldsymbol{\beta}}$ is the MLE for $\boldsymbol{\beta}$ but not the least squares estimate.**

Solution. Note that the above expression is minimized for $\boldsymbol{\beta}$ when $\beta = \hat{\boldsymbol{\beta}}$, in which case the second term equals zero (if $\beta \neq \hat{\boldsymbol{\beta}}$, then the second term is positive). Thus, $\hat{\boldsymbol{\beta}}$ is the MLE for $\boldsymbol{\beta}$. However, it is the least squares weighted estimate, which is not the same as the least squares estimate. ■

(c) **Define $\boldsymbol{Y}^* = (\boldsymbol{W}^{\frac{1}{2}})\boldsymbol{Y}$ and $\boldsymbol{X}^* = (\boldsymbol{W}^{\frac{1}{2}})\boldsymbol{X}$, where $\boldsymbol{W}^{\frac{1}{2}}$ is a symmetric square root matrix. Show this transforms the problem to the standard, unweighted regression model, and the usual Least Squares formula with $\boldsymbol{Y}^*$ and $\boldsymbol{X}^*$ implies the same value $\hat{\boldsymbol{\beta}}$.**

2

Solution. Note that with $\boldsymbol{Y}^* = (\boldsymbol{W}^{\frac{1}{2}})\boldsymbol{Y}$ and $\boldsymbol{X}^* = (\boldsymbol{W}^{\frac{1}{2}})\boldsymbol{X}$, we get that

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} = ((\boldsymbol{X}^*)^T\boldsymbol{X}^*)^{-1}(\boldsymbol{X}^*)^T\boldsymbol{Y}^*.$$

We can also use the associated $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$ values to find that

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})^T(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}).$$

Consequently, we have that the form matches the form of our standard unweighted regression model, and we get the formula

$$(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})^T(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) = (\boldsymbol{Y}^* - \boldsymbol{X}^*\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y}^* - \boldsymbol{X}^*\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\boldsymbol{X}^*)^T\boldsymbol{X}^*(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

which is also our typical least squares formula. This implies the same value for $\hat{\boldsymbol{\beta}}$, as shown above. ∎

(d) **Find the sampling distribution of $\hat{\boldsymbol{\beta}}$ (given $\boldsymbol{\beta}$ and σ^2).**

Solution. Note that $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}$ and $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{W}^{-1})$. It follows that

$$\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\mathbb{E}\left[\boldsymbol{Y}\right]$$
$$= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}$$
$$= \boldsymbol{\beta}.$$

Similarly,

$$\text{Var}\left[\hat{\boldsymbol{\beta}}\right] = \text{Var}\left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}\right]$$
$$= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\,\text{Var}\left[\boldsymbol{Y}\right]((\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W})^T$$
$$= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\sigma^2\boldsymbol{W}^{-1}((\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W})^T$$
$$= \sigma^2(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^{-1}\boldsymbol{W}^T\boldsymbol{X}((\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1})^T$$
$$= \sigma^2(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}((\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1})^T$$
$$= \sigma^2(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}.$$

Thus, we find that the sampling distribution for $\hat{\boldsymbol{\beta}}$ is

$$\boxed{\hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \sigma^2 \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1})}.$$ ∎

(e) **As an example where this would be useful, imagine a situation where the Normal linear model is appropriate, but where each independent response value Y_i is the average of m_i independent measurements. For example, the logs of the median selling prices in the home sales data are roughly the averages of individual log-selling prices for each municipality, with different numbers of houses sold (e.g. $m_i = 1767$ for Upper Darby and $m_i = 17$ for Swarthmore). Describe the appropriate $\boldsymbol{W}$ matrix.**

Solution. The $\boldsymbol{W}$ matrix in the home sales data would be a diagonal matrix (since the measurements are independent, we have no covariance in the non-diagonal entries) of size $\sum m_k$ over all municipalities. For the block of each matrix corresponding to a municipality of size m_k, the corresponding diagonal entries will be $f(m_k)$ for some function f (e.g. log).

The idea is that though the Normal linear model is appropriate, we put weights corresponding to the houses in each municipality, so that each house is equally weighted relative to the others in either the same or different municipalities. This ensures that we can weight the variances across each municipality, whereas in the Normal linear model we assume equal variance across all municipalities. ∎