JOE SEIDEL

# STAT 245
# HOMEWORK 5

*Projection Matrix 1.*

A symmetric Matrix $P \in \mathbb{R}^{n \times n}$ is a projection matrix if $P^2 = P$.

(a) Find $(I_n - P)^2$ and $(I_n - P)P$.

(b) Assume $\text{rank}(P) = r$, and then find all eigenvalues of $P$.

*Projection Matrix 2.*

A symmetric matrix $P \in \mathbb{R}^{n \times n}$ is a project matrix if $P^2 = P$.

(a) For any $X \in \mathbb{R}^{n \times p}$ such that $(X^T X)^{-1}$ exists. Show that $X(X^T X)^{-1} X^T$ is a projection matrix.

$$\begin{aligned}
(X(X^T X)^{-1} X^T)^2 &= X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T \\
&= X I_n (X^T X)^{-1} X^T \\
&= X(X^T X)^{-1} X^T
\end{aligned}$$

(b) Let $\mathbb{1} \in \mathbb{R}^{n \times 1}$ be a column vector of all ones. Show that $n^{-1}\mathbb{1}\mathbb{1}^T$ is a projection matrix.

$$\begin{aligned}
(n^{-1}\mathbb{1}\mathbb{1}^T)^2 &= n^{-1}\mathbb{1}\mathbb{1}^T n^{-1}\mathbb{1}\mathbb{1}^T \\
&= \frac{1}{n^2}\mathbb{1}\mathbb{1}^T\mathbb{1}\mathbb{1}^T \\
&= \frac{1}{n^2}\mathbb{1}n\mathbb{1}^T \\
&= n^{-1}\mathbb{1}\mathbb{1}^T
\end{aligned}$$

(c) If both $P_1$ and $P_2$ are projection matricies. Assume $P_1 P_2 = 0$. Show $P_1 + P_2$ is a projection matrix.

$$\begin{aligned}
(P_1 + P_2)^2 &= (P_1 + P_2)(P_1 + P_2) \\
&= P_1^2 + 2P_1 P_2 + P_2^2 \\
&= P_1^2 + P_2^2 \\
&= P_1 + P_2
\end{aligned}$$

*Projection Matrix 3.*

A symmetric matrix $P \in \mathbb{R}^{n \times n}$ is a projection matrix if $P^2 = P$. Assume $\text{rank}P = r$, for $z \sim N(0, I_n)$, use eigenvalue decomposition to find the distribution of $z^T P z$.

### *Variance Bias Trade-off.*

For any estimator $\hat{\theta}$, prove $\mathbb{E}(\hat{\theta}-\theta)^2 = \mathrm{Var}(\hat{\theta}) + (\mathbb{E}\hat{\theta}-\theta)^2$.

$$
\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= \mathbb{E}\left[(\hat{\theta}-\theta)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]+\mathbb{E}[\hat{\theta}]-\theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)^2 + 2\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)\left(\mathbb{E}[\hat{\theta}]-\theta\right) + \left(\mathbb{E}[\hat{\theta}]-\theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)^2\right] + \mathbb{E}\left[2\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)\left(\mathbb{E}[\hat{\theta}]-\theta\right)\right] + \mathbb{E}\left[\left(\mathbb{E}[\hat{\theta}]-\theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)^2\right] + 2\left(\mathbb{E}[\hat{\theta}]-\theta\right)\mathbb{E}\left[\hat{\theta}-\mathbb{E}[\hat{\theta}]\right] + \left(\mathbb{E}[\hat{\theta}]-\theta\right)^2 && \mathbb{E}[\hat{\theta}]-\theta = \text{const.} \\
&= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)^2\right] + 2\left(\mathbb{E}[\hat{\theta}]-\theta\right)\left(\mathbb{E}[\hat{\theta}]-\mathbb{E}[\hat{\theta}]\right) + \left(\mathbb{E}[\hat{\theta}]-\theta\right)^2 && \mathbb{E}[\hat{\theta}] = \text{const.} \\
&= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}[\hat{\theta}]\right)^2\right] + \left(\mathbb{E}[\hat{\theta}]-\theta\right)^2 \\
&= \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta},\theta)^2
\end{aligned}
$$

### *Variance Estimation 1.*

For the linear model $y \sim N(X\beta,\ \sigma^2 I_n)$., recal that $\hat{y} = Hy$ with $H = X(X^TX)^{-1}X^T$. The residual is defined as $\hat{e} = (I-H)y$. In the class we derived that $||\hat{e}||^2/\sigma^2 \sim \chi^2_{n-p}$. Use this fact to answer the following questions:

(a) Define $\hat{\sigma}^2 = \frac{1}{n-p}\|\hat{e}\|^2$. What is $\mathbb{E}\hat{\sigma}^2$?

(b) What is $\mathbb{E}(\hat{\sigma}^2-\sigma^2)^2$?

(c) Define $\tilde{\sigma}^2 = \frac{1}{n}\|\hat{e}^2\|^2$. What is $\mathbb{E}(\tilde{\sigma}^2-\sigma^2)$?

(d) Consider $\sigma_c^2 = c\|\hat{e}\|^2$. Find the $c$ such that $\mathbb{E}(\sigma_c^2-\sigma^2)^2$ is the smallest.

### *Variance estimation 2.*

Consider linear model $y \sim N(X\beta,\ ,\sigma^2 I_n)$.

(a) Find the joint MLE of $(\beta,\sigma^2)$, denoted as $(\hat{\beta},\hat{\sigma}^2)$.

(b) Construct a pivotal of $\sigma^2$ using $\hat{\sigma}^2$. Find an exact 95% confidence interval of $\sigma^2$.

### *Data Analysis*

Download NewHaven.txt from chalk. Set a up a working directory on your own computer, and read the data into R. Write a report of data analysis that addresses the following items. The report should be printed and submitted together with the homework. No need to include the code.

(a) Summarize the whole data set.

(b) Pick up a subset of rows that you want to study. For example you can study all houses or condos. You can also study houses that are not too expensive. Whatever subset of rows you pick, you need to justify your choice with some understandings of the data set.

(c) Pick up at least five variables. Explain what they are, and fit a linear models to predict current values.

(d) Analyze the inear model result. Use the function cooks.distance to find outliers. Use Google search to check why these are outliers. Fit the model after removing those outliders.

(e) You may want to repeat the last step. Write a nice paragraph with a clear conlcusion for your findings. Try to include many nice plots in your report that it is easy to read.