AMS256 Homework 1

1. Observe that $z_i = y_i/x_i = \beta_1 + \beta_0/x_i + \epsilon_i/x_i \Rightarrow$ This model is still linear in $\boldsymbol{\beta}$.

2. (a) Note that three columns are linearly independent. Or the three rows are linearly independent. So $r(\boldsymbol{X}) = 3$

   (b) $dim(C(\boldsymbol{X})) = r(\boldsymbol{X}) = 3$.
   A basis that we can easily come up with (you may find a different set of vectors for the basis) is a set of the following three vectors, $\boldsymbol{u}_1 = \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$ or $\boldsymbol{u}_2 = \begin{bmatrix}-1\\0\\0\\1\\0\end{bmatrix}$ or $\boldsymbol{u}_3 = \begin{bmatrix}1\\0\\0\\1\\0\end{bmatrix}$.

   (c) $dim(\boldsymbol{\mathcal{N}}(\boldsymbol{X})) = 4 - dim(C(\boldsymbol{X})) = 4 - 3 = 1$. We need to find a vector $\boldsymbol{v}$ such that $\boldsymbol{Xv} = \boldsymbol{0}$. A basis is $\boldsymbol{v}_1 = \begin{bmatrix}0\\1\\0\\-1\end{bmatrix}$. We can check that $\boldsymbol{v}_1 \perp \boldsymbol{u}_j$, $j = 1, 2, 3$.

   (d) $\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix}5 & 1 & 3 & 1\\1 & 3 & 1 & 3\\3 & 1 & 3 & 1\\1 & 3 & 1 & 3\end{bmatrix}$. Recognize that the submatrix of the upper 3×3 is nonsingular matrix and find a g-inverse matrix, $(\boldsymbol{X}^T\boldsymbol{X})^- = \begin{bmatrix}1/2 & 0 & -1/2 & 0\\0 & 3/8 & -1/8 & 0\\-1/2 & -1/8 & 7/8 & 0\\0 & 0 & 0 & 0\end{bmatrix}$

   (e)
$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T = \begin{bmatrix}1 & 0 & 0 & 0 & 0\\0 & 1/2 & 1/2 & 0 & 0\\0 & 1/2 & 1/2 & 0 & 0\\0 & 0 & 0 & 1/2 & 1/2\\0 & 0 & 0 & 1/2 & 1/2\end{bmatrix}$$
   If $\boldsymbol{y} \in C(\boldsymbol{P}) = C(\boldsymbol{X})$, $\boldsymbol{Py} = \boldsymbol{y}$. So $[3, 1, 1, 2, 2]^T$ and $[1, 0, 0, 2, 2]^T$ can be $\boldsymbol{Py}$.

   (f) Recall that $(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$ is a g-inverse of $\boldsymbol{X}$. We use this to construct all the solutions, $\tilde{\boldsymbol{x}} = \boldsymbol{Gc} + (I - \boldsymbol{GA})\boldsymbol{z}$ for $\boldsymbol{Ax} = \boldsymbol{c}$ from lecture. Thus, for some $\boldsymbol{z} \in \mathbb{R}^p$
$$\tilde{\boldsymbol{x}} = (\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y} + (I - (\boldsymbol{X}^T\boldsymbol{X})^-(\boldsymbol{X}^T\boldsymbol{X}))\boldsymbol{z}$$
   By varying $\boldsymbol{z}$, we can obtain all possible solutions of $\boldsymbol{\beta}$.

3.
$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \epsilon_i\\
&= \beta_0 + \beta_1(s + ti) + \epsilon_i\\
&= \beta_0 + \beta_1 s + \beta_1 ti + \epsilon_i\\
&= \gamma_0 + \gamma_1 i + \epsilon_i
\end{aligned}$$

where $\gamma_0 = \beta_0 + \beta_1 s$ and $\gamma_1 = \beta_1 t$. Thus, they are an equivalent parameterization.

4. • LSE

   Let $\hat{\theta} = argmin_\theta\ Q(\theta) = argmin_\theta\ ||\boldsymbol{Y} - \boldsymbol{X}\beta||$ where $\boldsymbol{X} = [x_1^2, \ldots, x_n^2]^T$ and $\beta = \theta$. From lecture, for full rank $\boldsymbol{X}$ we know

$$\hat{\beta} = \hat{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \frac{\sum_i x_i^2 y_i}{x_i^4}.$$

   • MLE

$$\frac{\partial log f(\mathbf{y}|\theta)}{\partial\theta} = \frac{\partial\{-\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_i (y_i - \theta x_i)^2)}{2\sigma^2}\}}{\partial\theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^4 x_i^4}$$

5. Plugging in $\boldsymbol{x}^\star$ and $\boldsymbol{y}^\star$ for $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, we obtain

$$\begin{aligned}
\hat{\beta_1}^\star &= \frac{S_{xy}^\star}{S_{xx}^\star} = \frac{\sum(x_i^\star - \bar{x}^\star)(y_i^\star - \bar{y}^\star)}{\sum(x_i^\star - \bar{x}^\star)^2} = \frac{\sum((c + dx_i) - (c + d\bar{x}))((a + by_i) - (a + b\bar{y}))}{\sum((c + dx_i) - (c + d\bar{x}))^2} \\
&= \frac{b\sum(x_i - \bar{x})(y_i - \bar{y})}{d\sum(x_i - \bar{x})^2} = \frac{b\hat{\beta}_1}{d}, \\
\hat{\beta_0}^\star &= \bar{y}^\star - \hat{\beta_1}^\star\bar{x}^\star = (a + b\bar{y}) - \frac{b\hat{\beta}_1}{d}(c + d\bar{x}) = a - \frac{bc\hat{\beta}_1}{d} + b\hat{\beta}_0.
\end{aligned}$$

6. From lecture, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$.

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 8 & 272 \\ 272 & 9318 \end{bmatrix} \quad \Rightarrow \quad (\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} 16.639 & -0.486 \\ -0.486 & 0.014 \end{bmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 71 & 63 & 68 & 70 & 71 & 63 & 68 & 70 \\ 2457 & 2010 & 2176 & 2176 & 2205 & 2412 & 2448 & 1920 \end{bmatrix}^T$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = \begin{bmatrix} 35.174 \\ 0.929 \end{bmatrix}$$

ANOVA Table:

Source of Variation	Sum of Squares	DF	Mean squares	F-stat
Model	60.415	1	60.415	50.73
Error(residual)	7.145	6	1.19	
Total	67.5	7	9.64	

$$\begin{aligned}
\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1\bar{x})^2 = \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\
&= \left(\frac{S_{xy}}{S_{xx}}\right)^2 \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{S_{xy}^2}{S_{xx}} = \boldsymbol{y}^T(\boldsymbol{P}_x - \boldsymbol{P}_1)\boldsymbol{y},
\end{aligned}$$

where $\boldsymbol{P}_x$ and $\boldsymbol{P}_1$ are the orthogonal projection operators onto $C(\boldsymbol{X})$ and $C(\mathbf{1})$, respectively. That is, $\boldsymbol{P}_x$: projection matrix for the model for an intercept and a slope and $\boldsymbol{P}_1$: projection matrix under the model with an intercept. From lecture,

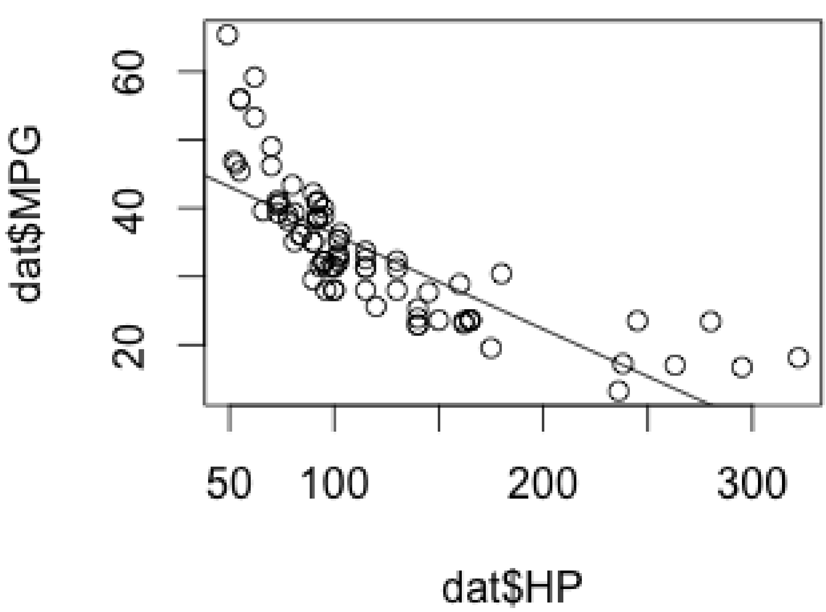
$$R^2 = 1 - \frac{||\boldsymbol{y}^T(\boldsymbol{P}_x - \boldsymbol{P}_1)\boldsymbol{y}||^2}{||\boldsymbol{y}^T(I - \boldsymbol{P}_1)\boldsymbol{y}||^2} = 1 - \frac{\sum_i(\hat{y}_i - \bar{y})^2}{\sum_i(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST} = 1 - \frac{7.145}{67.5} = 0.894$$

7. (a)

```
model1=lm(MPG~HP, dat)
plot(dat$HP, dat$MPG)
abline(model1)
summary(model1)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	50.0661	1.5695	31.90	0.0000
HP	-0.1390	0.0121	-11.52	0.0000

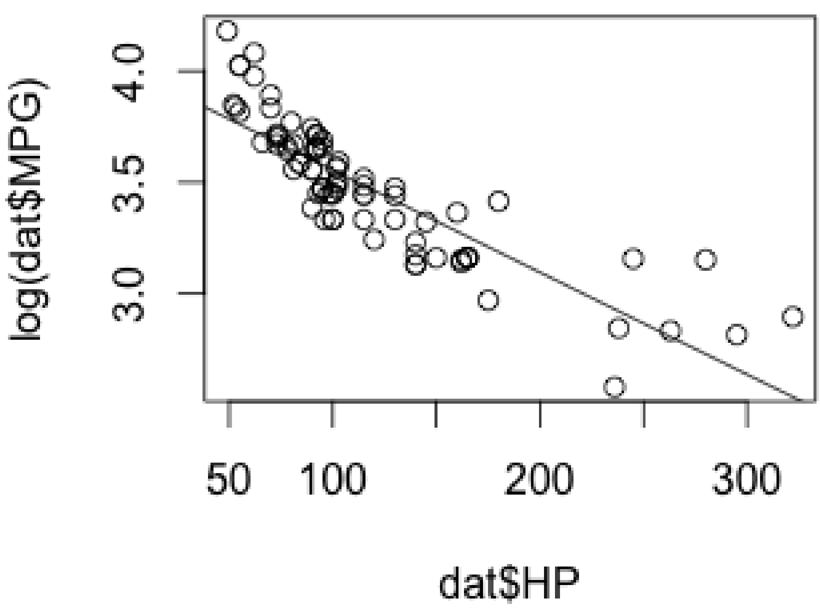
Residual standard error: 6.174 on 80 degrees of freedom
Multiple R-squared: 0.6239,
Adjusted R-squared: 0.6192
F-statistic: 132.7 on 1 and 80 DF,
p-value: $<$ 2.2e-16



(b)
```
model2=lm(log(MPG)~HP, dat)
plot(dat$HP, log(dat$MPG))
abline(model2)
summary(model2)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.0132	0.0401	100.02	0.0000
HP	-0.0046	0.0003	-14.87	0.0000

Residual standard error: 0.1578 on 80 degrees of freedom
Multiple R-squared: 0.7344,
Adjusted R-squared: 0.7311
F-statistic: 221.2 on 1 and 80 DF, p-value: $< 2.2e-16$



The scatterplots shows that the linear model fits better for log(MPG). Also, from the R output, Model2 has a higher Adjusted R-squared compared to Model1, which indicates that Model2 has better explains the variability in the data better than Model1. Model2 also have a lower residual standard error compared to Model1.

(c)
```
model3=lm(MPG~HP+VOL+SP+WT, dat)
summary(model3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	192.4378	23.5316	8.18	0.0000
HP	0.3922	0.0814	4.82	0.0000
VOL	-0.0156	0.0228	-0.69	0.4951
SP	-1.2948	0.2448	-5.29	0.0000
WT	-1.8598	0.2134	-8.72	0.0000

Residual standard error: 3.653 on 77 degrees of freedom
Multiple R-squared: 0.8733,
Adjusted R-squared: 0.8667
F-statistic: 132.7 on 4 and 77 DF, p-value: $<$ 2.2e-16

8.

$$
\begin{aligned}
\mathrm{E}(y_0) &= \mathrm{E}(\hat{\beta}_0 + \hat{\beta}_1 x_0 + \epsilon_0) \\
&= \beta_0 + \beta_1 x_0
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(y_0) &= \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_0 + \epsilon_0) \\
&= \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_0) + \mathrm{Var}(\epsilon_0) \\
&= \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_0) + \sigma^2 \\
&= \begin{bmatrix} 1 & x_0 \end{bmatrix} \begin{bmatrix} \frac{\sigma^2}{nS_{xx}} \sum x_i^2 & -\frac{\sigma^2 \bar{x}}{nS_{xx}} \\ -\frac{\sigma^2 \bar{x}}{nS_{xx}} & \frac{\sigma^2}{nS_{xx}} \end{bmatrix} \begin{bmatrix} 1 \\ x_0 \end{bmatrix} + \sigma^2 \\
&= \frac{\sigma^2}{nS_{xx}} \left(\frac{\sum x_i^2}{n} - 2\bar{x} x_0 + x_0^2 \right) + \sigma^2
\end{aligned}
$$