Statistics [Q3] [30 marks]

a) **[9 marks]**

i) **[2 marks]**
$\hat{\pi} = (0.067 + 0.017 + 0.017) \times 1.5 = 0.1515$,

ii) **[4 marks]**
$z_{0.98} = 2.054$ **[1 mark]**, so
standard error $= z_{0.98}\sqrt{\hat{\pi}(1-\hat{\pi})/n} = 0.0823$,
giving a 96% CI for π of $\hat{\pi} \pm z_{0.98}\sqrt{\hat{\pi}(1-\hat{\pi})/n} = (0.0692, 0.2338)$ **[3 marks: 1 mark for the correct equation, 2 marks for the correct interval]**.

iii) **[3 marks]**
CLT empirical rule $n\hat{\pi}(1-\hat{\pi}) = 80 * 0.1515 * (1 - 0.1515) = 10.2838 > 5$ **[2 marks]**
So, the sample size is large enough. **[1 mark]**

b) **[21 marks]**

i) **[1 mark]** $r^2 = 0.941$

ii) **[1 mark]** $s = 0.312$

iii) **[7 marks: 2 marks for hypotheses, 1 mark for df, 2 marks for rejection region (or 1 mark for observed test statistic, 1 mark for p-value), 2 mark for conclusion]**

- $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ [2 marks: 1 mark for H_0 and 1 mark for H_a]
- $t_{13,0.975} = 2.160$.
  Rejection criterion: Reject H_0 if

$$\hat{\beta}_1 \notin \left[-t_{13,0.975}\frac{s}{\sqrt{s_{xx}}}, t_{13,0.975}\frac{s}{\sqrt{s_{xx}}}\right] = [-2.160 \times 0.011475, 2.160 \times 0.011475] = [-0.0248, 0.0248]\,.$$

  Or, $t = \dfrac{-0.16506}{0.011475} = -14.3843$, $df = 13$. $p-value = 2P(T > 14.3843) = 2.3125e{-}9$.
- Reject H_0. `Time` is significantly associated with `Resistance`. (or something similar that ties original problem with statistical results.)

iv) **[4 marks: 1 mark for $\hat{y}(x_0)$, 1 mark for t quantile, 1 mark for correct equation, 1 mark for correct interval]**
$\hat{y}(12) = 5.5931 - 0.16506(12) = 3.6124$, $t_{13,0.95} = 1.771$

$$\begin{aligned} & \hat{y}(x_0) \pm st_{n-2,1-\alpha/2}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}}} \\ = \; & 3.6124 \pm 0.312 \times 1.771\sqrt{\frac{1}{15} + \frac{(12 - 272/15)^2}{739.733}} \\ = \; & [3.4230, 3.8018] \end{aligned}$$

v) **[4 marks: 1 mark for $\hat{y}(x_0)$, 1 mark for t quantile, 1 mark for correct equation, 1 mark for correct interval]**
$\hat{y}(16) = 5.5931 - 0.16506(16) = 2.9521$, $t_{13,0.99} = 2.650$

$$\begin{aligned} & \hat{y}(x_0) \pm s t_{n-2,1-\alpha/2} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}}} \\ = \; & 2.9521 \pm 0.312 \times 2.650 \sqrt{1 + \frac{1}{15} + \frac{(16 - 272/15)^2}{739.733}} \\ = \; & [2.0957, 3.8085] \end{aligned}$$

vi) **[4 marks]**

i. **[3 marks: 1 mark per correct assumption]**
- $e_i's$ have been drawn independently of one another
- $e_i's$ have the same variance
- $e_i's$ have been drawn from a normal distribution

ii. Residual plot does not show obvious pattern. (If student wrote "there is a quadratic pattern", it is still fine.) **[0.5 mark]**
QQ plot is closed to a straight line. **[0.5 mark]**

Statistics [Q4] [30 marks]

a) **[3 marks: 1 mark for each point]**

- *Comment about location:* The apparent ordering of fuel efficiency (in MPG, from best to worst) is FWD, RWD and AWD.
- *Comment about spread:* More variability for FWD than the others.
- *Comment about shape/outliers:* Right skewed for FWD and roughly symmetric for the others. Many outliers for both FWD and AWD.

b) **[4 marks]**

- The observations for fuel efficiency by drivetrain were drawn from Normal distributions. **[1 mark]**
- The observations are independent. **[1 mark]**
- The variances for fuel efficiency by drivetrain are the same. **[1 mark]**
  This assumption can be checked given the summary statistics. Using the rule-of-thumb (i.e., the ratio of the largest sample standard deviation to the smallest one is smaller than 2), this assumption is NOT acceptable. **[1 mark]**

c) **[5 marks: 1 mark for each missing value in the table]**

Source	df	SS	MS	F
Treatment	2	2885	1442.5	59.08
Error	384	9376.1	24.42	
Total	386	12261.1		

d) **[7 marks: 2 marks for hypotheses, 1 mark for df, 1 mark for observed test statistic, 1 mark for rejection region (or 1 mark for p-value), 2 mark for conclusion]**

- $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_a : not all the means are equal (an alternative hypothesis stated as $H_a : \mu_1 \neq \mu_2 \neq \mu_3$ is not correct).
- The observed value of the test statistic is $f_0 = 59.08$.
- Rejection criterion: reject H_0 if $f_0 > f_{2,21;0.95} = 3.0192$.
  Or, the p-value is $p = P(F_{2,384} > 59.08)$. $p < 0.001$.
- Conclusion: Reject H_0. There is very strong evidence that fuel efficiency among drivetrains are not all the same.

e) **[4 marks: 1 mark for the t critical value, 1 mark for a correct expression of the CI, 2 marks for the correct values]**

$$\left[(\bar{x}_1 - \bar{x}_3) \pm t_{n-k;1-\alpha/2}\sqrt{\mathrm{MS}_{\mathrm{Er}}\left(\frac{1}{n_1} + \frac{1}{n_3}\right)}\right]$$
$$= \left[(29.60 - 25.46) \pm 1.650\sqrt{24.42 \times \left(\frac{1}{215} + \frac{1}{94}\right)}\right]$$
$$= [3.5290, 4.7510].$$

f) **[7 marks: 2 marks for hypotheses, 1 mark for df, 1 mark for observed test statistic, 1 mark for p-value, 1 mark for mentioning (or effectively using) a Bonferonni adjustment, 1 mark for a correct conclusion]**

- $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$
- Observed value of the test statistic

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\mathrm{MS}_{\mathrm{Er}}(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{29.60 - 23.01}{\sqrt{24.42 \times (\frac{1}{215} + \frac{1}{78})}} = 10.0889$$

  p-value is $p = 2 \times \mathbb{P}(T > 10.0889)$ for $T \sim t_{384}$. $p < 0.0001$.
- p-value $< \alpha/3 = 0.0167$ (Bonferonni adjustment). So, reject H_0, and the conclusion is the same as (d).