Instituto Superior Técnico — Departamento de Matemática

Linear Model Analysis (LMAC/MECD/MMAC)

Exam — 1st semester – 2023/24

Duration: 2 hours — 17/01/24 – 18h

Please explain your answers appropriately!

The effectiveness of a new experimental overdrive gear in reducing gasoline consumption was studied in 12 trials with a light truck equipped with this gear. In the data that follows, X_i denotes the constant speed (in miles per hour) on the test track in the ith trial and Y_i denotes miles per gallon obtained.

i	1	2	3	4	5	6	7	8	9	10	11	12
X_i	35	35	40	40	45	45	50	50	55	55	60	60
Y_i	22	20	28	31	37	38	41	39	34	37	27	30

The researcher in charge of the study proposed three candidate models, using the centered version of X, $x_i = X_i - \overline{X}$, to analyze this data set.

Model I: $Y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i^I$, $i = 1, \ldots, 12$, $\varepsilon_i^I \overset{iid}{\sim} N(0, \sigma_I^2)$

Model II: $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i^{II}$, $i = 1, \ldots, 12$, $\varepsilon_i^{II} \overset{iid}{\sim} N(0, \sigma_{II}^2)$

Model III: $Y_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \gamma_3 x_i^3 + \varepsilon_i^{III}$, $i = 1, \ldots, 12$, $\varepsilon_i^{III} \overset{iid}{\sim} N(0, \sigma_{III}^2)$

The models were fit in R and the following outputs were obtained.

Model I:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.0000     1.8572  17.231 9.17e-09 ***
x             0.3314     0.2175   1.524    0.159
---
Residual standard error: 6.433 on 10 degrees of freedom
Multiple R-squared:  0.1885,    Adjusted R-squared:  0.1073
F-statistic: 2.322 on 1 and 10 DF,  p-value: 0.1585
```

Model II:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.640625   0.766888  50.386  2.4e-12 ***
x            0.331429   0.058372   5.678 0.000303 ***
I(x^2)      -0.091071   0.007993 -11.394  1.2e-06 ***
---
Residual standard error: 1.727 on 9 degrees of freedom
Multiple R-squared:  0.9474,    Adjusted R-squared:  0.9357
F-statistic: 81.03 on 2 and 9 DF,  p-value: 1.757e-06
```

Model III:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.640625   0.777192  49.718 2.97e-11 ***
x            0.467030   0.166135   2.811   0.0228 *
I(x^2)      -0.091071   0.008100 -11.243 3.52e-06 ***
I(x^3)      -0.001074   0.001230  -0.873   0.4079
---
Residual standard error: 1.75 on 8 degrees of freedom
Multiple R-squared:  0.952,     Adjusted R-squared:  0.934
F-statistic: 52.85 on 3 and 8 DF,  p-value: 1.284e-05
```

Some useful sums:

```
sum(Y)  sum(Y^2)
  384     12798

sum(X)   sum(X^2)   sum(X^3)   sum(X^4)
  570      27950     1410750   73043750

sum(x)   sum(x^2)   sum(x^3)   sum(x^4)
  0.0      875.0        0.0    110468.8
```

1. The researcher concluded that the best model is Model II. Do you agree? Why? Present at least one criterion on which Model II is the best. (2.0)

2. Present the ANOVA table for Model II and conclude whether or not the regression is significant. Use $\alpha = 0.05$. (2.0)

3. Express the fitted regression equation for Model II in terms of the original variable X. Comment. (2.0)

4. Calculate the coefficient of correlation between X and X^2 and between x and x^2. Is the use of a centered variable helpful here? (2.0)

The researcher used the predict function in R and Model II to produce the following table.

X_0	x_0	$\hat{\mu}_{y\|x_0}$	$s.e.\left(\hat{\mu}_{y\|x_0}\right)$
35	-12.5	20.27	1.11
40	-7.5	31.03	0.68
45	-2.5	37.24	0.74
50	2.5	38.90	0.74
55	7.5	36.00	0.68
60	12.5	28.55	1.11

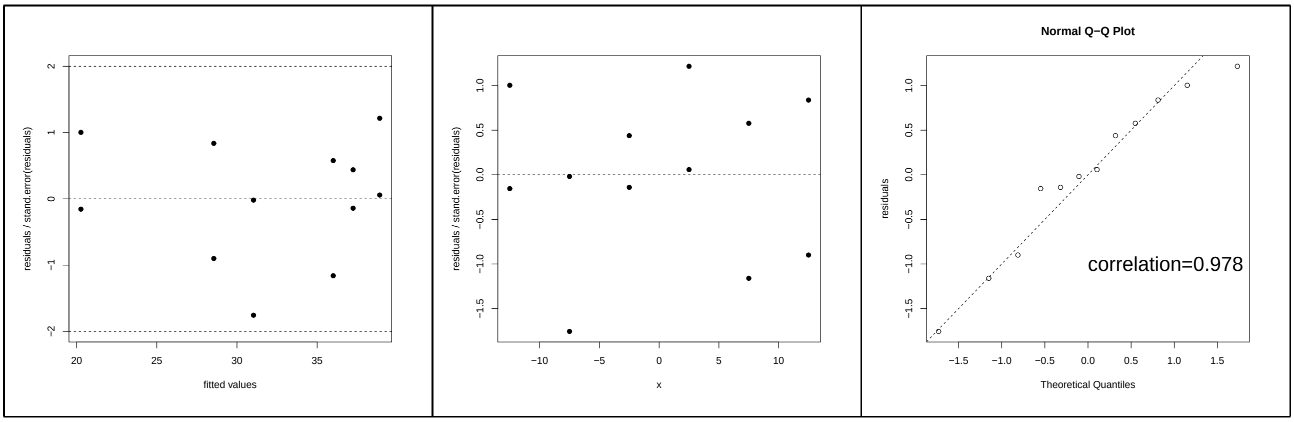
You may use these results to answer the next questions.

5. Assuming Model II, obtain joint interval estimates for the mean miles per gallon for tests run at speeds 40 and 50 miles per hour. Use the most efficient simultaneous estimation procedure and a 90% family confidence coefficient. Interpret the intervals. (2.0)

6. Assuming Model II, compute an interval, which contains with 90% confidence the miles per gallon in the next test run at 50 miles per hour. How is this interval called? Can you compare it with any of the intervals computed in question 5? (2.0)

7. Test whether the quadratic term can be dropped from Model II. Use $\alpha = 0.05$. (2.0)

8. The researcher also produced the following diagnostic plots for Model II. What can you conclude from each plot? (2.0)



9. Using the additional information that $SSPE = 18$, test Model II for lack-of-fit. (2.0)

10. Now consider Model III, and test whether or not $\gamma_3 = 0$ using $\alpha = 0.05$. How can your conclusion be related with the result from the test in question 9? (2.0)