(a) Summarize the conclusions that should be drawn from the analysis represented by the listing. 6 Marks

(b) Summarize the conclusions that should be drawn from the residual plot. 2 Marks

(c) What is the predicted `IGG` level for a study subject with `MAXOXY` level 50 ? Justify your answer. 2 Marks

(d) The $(1 - 0.025) = 0.975$ quantile of the Student-t distribution with 28 degrees of freedom is 2.0484. Use this information, in conjunction with the information in the listing, to construct 95% confidence intervals for parameters β_0 and β_1. 4 Marks

(e) A $(1 - \alpha) \times 100\%$ prediction interval for the predicted response at $x = x^{\text{new}}$ takes the form

$$\widehat{y}^{\text{new}} \pm t_{\alpha/2,n-2} \times \sqrt{\widehat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x^{\text{new}} - \overline{x})^2}{S_{xx}}\right)}$$

where

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2.$$

In the study data, the mean of the `MAXOXY` measurements is 50.636 units, and the quantity S_{xx} is equal to 4171.37 (units squared).

Is the prediction interval wider for a prediction made at $x^{\text{new}} = 55$ or $x^{\text{new}} = 40$? Justify your answer. 2 Marks

(f) Explain the difference between a *prediction interval* and a *confidence interval* for a prediction from the simple linear regression. 4 Marks

Answer to Q2:

(a) From lines 15 and 21, we see that there is a significant relationship between predictor and response. On line 20, the R^2 statistic is quite high, indicating that the predictive capability of the model is good. The intercept is not significantly different from zero (line 16), which is intuitively reasonable in this setting. The relationship between x and y is therefore approximately

$$y = 32x$$

as the intercept is not significantly different from zero.

8 Marks

(b) The residual plot seems to indicate that the model fit is adequate, as there is no pattern, and a constant variability around zero.

2 Marks

(c) Prediction is $\widehat{y} = -100.345 + 32.743 \times 50 = 1536.805$.

2 Marks

(d) The two confidence intervals are

$$\beta_0 \ : \ \widehat{\beta}_0 \pm 2.0484 \times \text{e.s.e}(\widehat{\beta}_0) = -100.345 \pm 2.0484 \times 100.450 = (-306.1068 : 105.4168)$$

and

$$\beta_1 \ : \ \widehat{\beta}_1 \pm 2.0484 \times \text{e.s.e}(\widehat{\beta}_1) = 32.743 \pm 2.0484 \times 1.932 = (28.7855 : 36.7005)$$

4 Marks

(e) The prediction interval width is determined as a monotonic function of $(x^{\text{new}} - \overline{x})^2$, so is wider when x^{new} is further away from $\overline{x}$; here the value $x^{\text{new}} = 55$ is closer to the sample mean, so that gives the narrower interval.

2 Marks

(f) A **prediction interval** at $x = x_1^{\text{new}}$ incorporates the random variation that is present in the observations into interval, whereas the **confidence interval** does not. We write

$$\widehat{Y}_{\text{O}}^{\text{new}} = \widehat{Y}^{\text{new}} + \epsilon^{\text{new}}$$

where ϵ^{new} is a zero mean, variance σ^2 random residual error, independent of all other random quantities. Then

$$\begin{aligned}
\text{Var}_{\mathbf{Y}|\mathbf{X}}[\widehat{Y}_{\text{O}}^{\text{new}}|\mathbf{X}] &= \text{Var}_{\mathbf{Y}|\mathbf{X}}[\widehat{Y}^{\text{new}}|\mathbf{X}] + \text{Var}_{\mathbf{Y}|\mathbf{X}}[\epsilon^{\text{new}}|\mathbf{X}] \\
&= \sigma^2 h^{\text{new}} + \sigma^2 \\
&= \sigma^2(1 + h^{\text{new}}).
\end{aligned}$$

Thus the prediction interval is **wider** than the confidence interval.

4 Marks

Answer to Q :

Answer to Q (continued):