Degree Exam: Level M Regression Modelling

1 (a)

(i) $$\underline{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix}; \quad X = \begin{array}{l} n_1\left\{ \vphantom{\begin{matrix}1\\1\\1\end{matrix}} \right. \\ n_2\left\{ \vphantom{\begin{matrix}1\\1\\1\end{matrix}} \right. \end{array} \begin{pmatrix} 1 & 1 & 0 \\ | & | & | \\ | & 1 & 0 \\ | & 0 & 1 \\ | & | & | \\ 1 & 0 & 1 \end{pmatrix}; \quad \underline{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

(ii) $$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_1 - \bar{x} \\ | & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}; \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

(iii) $$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{2n} \end{pmatrix}; \quad X = n\left\{ \begin{pmatrix} 1 & x_1 & 1 & x_1 \\ | & \vdots & \vdots & \vdots \\ | & \vdots & 1 & x_n \\ | & \vdots & 0 & 0 \\ | & \vdots & \vdots & \vdots \\ 1 & x_{2n} & 0 & 0 \end{pmatrix} \right.; \quad \underline{\beta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

(b) Residual: $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$

Studentized residual: $r_i^* = \dfrac{Y_i - \hat{Y}_i}{\sqrt{\sigma^2(1-h_{ii})}}$

where $h_{ii} = (i,i)$th element of $H = X(X^TX)^{-1}X^T$.

(c) $$\begin{aligned} \underline{\hat{\varepsilon}} &= \underline{Y} - \underline{\hat{Y}} \\ &= \underline{Y} - X\underline{\hat{\beta}} \\ &= \underline{Y} - \underbrace{X(X^TX)^{-1}X^T}_{=}\underline{Y} \\ &= (I - H)\underline{Y} \end{aligned}$$

so each $\hat{\varepsilon}_i$ is a linear combination of the Y_is.

(d) Consider the linear model $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$. Assume

$$E(\hat{\underline{\varepsilon}}) = \underline{0}$$

and $\text{cov}(\hat{\underline{\varepsilon}}) = \sigma^2 I.$

Then among all unbiased linear estimators of $\underline{\beta}$ the least squares estimator (LSE) has minimum variance.

(e) $A = \underline{Y}$; $B = \hat{\underline{\varepsilon}} = \underline{Y} - \hat{\underline{Y}}$; $C = \hat{\underline{Y}}$.

② (a) Constant variance: graph of residuals versus fitted values. if it fans out then no constant variance.

Normality of errors: Q-Q plot; line should be aligne line.

or graph of studentized residuals (r_i^*, which should be normally distributed) versus fitted values; if there are a lot of observations with $|r_i^*| > 3$, then we found evidence that the errors are not normally distributed.

(b) Leverage of ith observation is h_{ii}, the (i,i)th element of the hat matrix $H = X(X^TX)^{-1}X^T$.

Large h_{ii} means the point is far apart from the centroid (it might be influential or not).

(c)

Any two models ⟨ minimum Mean RSS or minimum AIC

With intercept ⟨ same number of explanatory variables: R^2 ; different " " " " : R_a^2

Without intercept ⟨ same " " " " : R_0^2 ; different " " " " : $R_{0,a}^2$

③ a)

(i) $a = \dfrac{32.724}{1} = 32.724$;

$$b = \frac{32.724}{0.702} = 46615.38$$

$$c = 43.255 - 32.724 = 10.531$$

$$R^2 = 1 - \frac{RSS_0}{TSS} = 1 - \frac{10.531}{43.255} = 0.7565$$

Temperature explains 75.7% of the variability of the number of chirps so it is a high percentage.

(ii) $\underline{b}^T \underline{\hat{\beta}} \pm t(n-2, 0.975) \sqrt{\dfrac{RSS}{n-k} \underline{b}^T (X^T X)^{-1} \underline{b}}$

$\underline{b}^T = (0,1)$

$$(0,1)\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \pm \underbrace{t(15, 0.975)}_{2.131} \sqrt{\underbrace{\frac{10.531}{17-2}(0.00416383)}_{0.002921568}}$$

$$0.37 \pm 2.131(0.05405153)$$

$$= (0.255, 0.485)$$

It does not contain 0 so we found evidence that the no. of chirps depends on the temperature

(4) (i) δ i.e. the interaction between concentration and temperature.

No it is not.

(ii) $R^2 = 0.54$ not very high. Concent. and temperature explain 54 % of the variability of the percentage of shrinkage.

(iii) Not adequate as interaction is not significant

Model (2) is the parallel lines model. It is adequate as both concentration and temperature are significantly related to the percentage of shrinkage so pick this model.

(iv)

Model (1) (with the interaction)

significant interaction → End and pick model (1)

not sign. int. → Model (2) (without int., with 2 main effects)

both main effects sign. → End and pick model (2)

only 1 main effect is significant → Remove the non-sign. effect and fit resulting model

remaining effect sign. → End

not sign. → Fit null model. End

none of the effects sign. → We can remove one of the and fit the model. Repeat with the other effect. Etc, Keep only significant effects

Wald Test statistic of the parameter that we are testing.

(iv) Some heteroskedasticity as the scatter ~~spread~~ ~~off~~ of the points ~~is~~ has smaller spread close to the middle. Maybe a non-linear model would be better as there is a little bit a non-linear pattern: