EPFL | MATH-408
Autumn Semester | 2025–2026

Regression Methods: Mock Exam Solutions

18 November 2025

Exercise 1 See lecture notes

Exercise 2

1. $\Pr(Z_j = 1) = 1 - F\{-x_j^T(\gamma/\sigma)\}$, therefore $\beta = \gamma/\sigma$ is estimable, but γ and σ are unidentifiable. Knowing Z can only tell us about the probability that $Y > 0$, but it cannot tell us how spread out the distribution of Y is, i.e., it cannot tell us σ.

2. The density of Z_j can be written in terms of $\pi_j = \Pr(Z_j = 1)$ as $\pi_j^{z_j}(1-\pi_j)^{1-z_j}$ and the Zs are independent, so the likelihood is

$$\prod_{j=1}^{n} \pi_j^{z_j}(1-\pi_j)^{1-z_j}.$$

Now $\mathrm{E}(Z_j) = \mu_j = \pi_j(-\eta_j) = 1 - F(-\eta_j)$, where $\eta_j = x_j\beta$ is the linear predictor. The response distribution for Z is binary and the link function is given by $\eta = g(\mu) = -F^{-1}(1-\mu)$.

In both cases (i) and (ii) the response distribution is binary (obviously).

In (i), F is the standard normal distribution Φ, so the link function is $g(\mu) = -\Phi^{-1}(1-\mu) = \Phi^{-1}(\mu)$, which is the probit link function.

In (ii), $F(\eta) = \exp\{-\exp(-\eta)\}$, so $g(\mu) = -F^{-1}(1-\mu) = \log\{-\log(1-\mu)\}$, which is the complementary log-log link function.

3. If we have n independent individuals whose responses $I_1, \dots, I_K$ fall into the set $\{1, \dots, K\}$, corresponding to K ordered categories, and that

$$\gamma_l = \mathrm{P}\left(I_j \le l\right) = \pi_1 + \cdots + \pi_l, \quad l = 1, \dots, K, \quad \gamma_K = 1,$$

this is an ordinal odds model, useful when there are several ordered categories (i.e., curry is mild, medium spicy, volcanic) and each individual (curry) is classified into one of them. There may be other variables x (e.g., the cook, the restaurant, tandoori, tikka, balti, …) that can influence the category. The likelihood is is

$$\prod_{j=1}^{n}\prod_{k=1}^{K} \Pr(Y_j \in \mathcal{I}_k) = \prod_{j=1}^{n}\prod_{k=1}^{K} \pi_k^{I(Y_j \in \mathcal{I}_k)}$$

where

$$\pi_l = \mathrm{P}\left(\zeta_{l-1} < x_j^{\mathrm{T}}\beta + \sigma\varepsilon \le \zeta_l\right) = F\left(\frac{\zeta_l - x_j^{\mathrm{T}}\beta}{\sigma}\right) - F\left(\frac{\zeta_{l-1} - x_j^{\mathrm{T}}\beta}{\sigma}\right), \quad l = 1, \dots, K.$$

Here we see that if we map $\zeta_1, \dots, \zeta_{K-1} \mapsto \zeta_1 + a, \dots, \zeta_{K-1} + a$ and $\beta_0 \mapsto \beta_0 - a$, then the model is unchanged for any a, so we must set $\beta_0 = 0$ or fix one of the ζs to get an estimable model. In this case σ and γ_1 can both be estimated, as having $K > 2$ categories will provide information about the spread of ϵ.

Exercise 3

1. A hat matrix $H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ is obviously symmetric and idempotent.

2. The subspaces satisfy $\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2$ because $X = (X_0, X_1, X_2)$, and therefore $H_0 y \in \mathcal{V}_0$ also lies in $\mathcal{V}_1$, which implies that $H_1 H_0 y = H_0 y$, for any $y \in \mathbb{R}^n$. Hence $H_1 H_0 = H_0$, which in turn implies that $(H_1 H_0)^T = H_0^T H_1^T = H_0 H_1 = H_0^T = H_0$, as required. The corresponding equations sought are $H_0 H_2 = H_2 H_0 = H_0$ and $H_1 H_2 = H_2 H_1 = H_1$. This implies that

   $$(H_1 - H_0)(H_2 - H_1) = H_1 H_2 - H_0 H_2 - H_1 H_1 + H_0 H_1 = H_1 - H_0 - H_1 + H_0 = 0,$$

   with other similar computations giving that $H_1 - H_0$, $H_2 - H_1$, H_0 and $(I_n - H_2)$ are all orthogonal. If $y \sim \mathcal{N}_n(\mu, \sigma^2 I_n)$ the vectors on the right-hand side of the decomposition are orthogonal and therefore are independent.

3. The full matrix will have a column of ones corresponding to μ, four columns corresponding to the treatments and eight columns corresponding to the subjects, but one of those for the treatments and one of those for the subjects must be dropped in order for X to be of full rank. Thus we have $X_0 = 1_n$, X_1 of size $n \times 3$ and X_2 of size $n \times 7$. The sums of squares are then respectively $\|(H_1 - H_0)y\|^2 = y^{\mathrm{T}}(H_1 - H_0)y$ (for treatments), $\|(H_2 - H_1)y\|^2 = y^{\mathrm{T}}(H_2 - H_1)y$ (for subjects, after allowing for treatments) and $\|(I_n - H_2)y\|^2 = y^{\mathrm{T}}(I_n - H_2)y$ (the residual after allowing for subjects and treatments), while the total is $\|(I_n - H_0)y\|^2 = y^{\mathrm{T}}(I_n - H_0)y$, where $n = 32$.

   The numbers in the third column are the respective ranks of the matrices $H_1 - H_0$, $H_1 - H_2$, $I_n - H_2$ and $I_n - H_0$; $d_{\mathrm{S}} = 7$ and $d_{\mathrm{Tot}} = 31$

Exercise 4

1. If $\beta = \beta'$, then $\hat{\beta} \sim \mathcal{N}_p\{\beta', \sigma^2(X^{\mathrm{T}}X)^{-1}\}$ and therefore $(\hat{\beta} - \beta')^{\mathrm{T}} X^{\mathrm{T}} X (\hat{\beta} - \beta')/\sigma^2 \sim \chi_p^2$, independent of the residual sum of squares. If σ^2 is unknown, then it can be estimated by $s^2 = y^{\mathrm{T}}(I - H)y/(n - p)$, and then under the null hypothesis we have

   $$F = \frac{(\hat{\beta} - \beta')^{\mathrm{T}} X^{\mathrm{T}} X (\hat{\beta} - \beta')/p}{s^2} \sim F_{p.n-p}.$$

2. Although there are three angles, with angles α, β, γ, say, their sum is the constant $\alpha + \beta + \gamma = \pi$, and so just two angles can vary independently. In terms of α and β, we have $y_A = \alpha + \varepsilon_A$, $y_B = \beta + \varepsilon_B$, and $y_C = \pi - \alpha - \beta + \varepsilon_C$, and this gives the linear model

   $$\begin{pmatrix} y_A \\ y_B \\ y_C - \pi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_A \\ \varepsilon_B \\ \varepsilon_C \end{pmatrix}.$$

   Hence

   $$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \pi + y_A - y_C \\ \pi + y_B - y_C \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2y_A + \pi - y_B - y_C \\ 2y_B + \pi - y_A - y_C \end{pmatrix}$$

   It is straightforward to show that $s^2 = (y_A + y_B + y_C - \pi)^2/3$.

   The triangle is equilateral if $\alpha = \beta = \pi/3$, which corresponds to the setup in (1.) with $\beta' = (\pi/3, \pi/3)^{\mathrm{T}}$, and would lead to a test based on an $F_{2,1}$ statistic.