EXERCISE 03/09/2024

a) Consider a model $Y_i = \mu_i + \varepsilon_i \quad i = 1, \dots, n$

The assumptions of a Gaussian linear model are:

I. normality, homoscedasticity, independence : $\varepsilon_i \sim N(0, \sigma^2)$ iid $i = 1, \dots, n$

II. linearity : $\mu_i = \underline{\tilde{x}}_i^T \underline{\beta}$ (linear w.r.t. $\underline{\beta}$) $i = 1, \dots, n$

(III. linearly independent covariates.)

1. yes, all assumptions are satisfied.
2. no, it is not linear in $\underline{\beta}$ and it can not be transformed into a linear model
3. $\log(Y_i) = \beta_1 \frac{x_{i1}}{x_{i2}} + \beta_3 \cdot \frac{\log x_{i3}}{x_{i2}} + \varepsilon_i$ yes
4. no, but the log transformation is

$$\log(Y_i) = \log(\beta_1) + \log(x_{i2}^{\beta_2}) + \log(\exp\{\varepsilon_i\})$$

$$\Rightarrow \underbrace{\log(Y_i)}_{=Y_i^*} = \underbrace{\log \beta_1}_{=\beta_1^*} + \underbrace{\beta_2}_{=\beta_2^*} \underbrace{\log x_{i2}}_{=x_i^*} + \varepsilon_i \qquad \varepsilon_i \sim N(0,1) \text{ indep.}$$

consider the transformed model

$$Y_i^* = \beta_1^* + \beta_2^* x_{i2}^* + \varepsilon_i$$

with

$$Y_i^* = \log Y_i$$
$$\beta_1^* = \log \beta_1 \qquad \beta_2^* = \beta_2 \qquad \Rightarrow \text{ it satisfies all assumptions}$$
$$x_i^* = \log x_{i2}$$

b) $\underline{Y}^*$ is a vector of random variables of dimension n

$$\underline{Y}^* = [Y_1^*, \dots, Y_n^*]^T$$

X^* is a $(n \times 2)$ matrix of known constants

$$X^* = \begin{bmatrix} 1 & \log x_1 \\ \vdots & \vdots \\ 1 & \log x_i \\ \vdots & \vdots \\ 1 & \log x_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12}^* \\ \vdots & \vdots \\ 1 & x_{i2}^* \\ \vdots & \vdots \\ 1 & x_{n2}^* \end{bmatrix}$$

$\underline{\beta}^*$ is a 2-dim vector of unknown constants

$$\underline{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} = \begin{bmatrix} \log \beta_1 \\ \beta_2 \end{bmatrix}$$

$\underline{\varepsilon}^*$ is an n-dim vector of random variables

$$\underline{\varepsilon}^* \sim N_n(\underline{0}, I_n) \qquad I_n \text{ identity matrix } (n \times n)$$

c) There is only one covariate → simple lm.

Hence, we know that the estimates are:

$$\hat{\beta}_1^* = \bar{y}^* - \hat{\beta}_2^* \bar{x}^*$$

$$\hat{\beta}_2^* = \frac{\text{cov}(x^*, y^*)}{\text{var}(x^*)} = \frac{\sum_{i=1}^n (x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2}$$

or, equivalently, $\underline{\hat{\beta}}^* = (X^{*T} X^*)^{-1} X^{*T} \underline{y}^*$

The estimator $\underline{\hat{B}}^*$ hence is

$$\underline{\hat{B}}^* = \begin{bmatrix} \bar{Y}^* - \hat{B}_2^* \bar{x}^* \\ \dfrac{\sum_{i=1}^n (x_i^* - \bar{x}^*)(Y_i^* - \bar{Y}^*)}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2} \end{bmatrix}$$

$$\underline{\hat{B}}^* \sim N_2(\underline{B}^*, \text{var}(\hat{B}^*))$$
$$\text{with} \quad \text{var}(\hat{B}^*) = (X^{*T} X^*)^{-1}$$

d) $\sum_{i=1}^n e_i = 0$ yes,

The sum of the residuals in this case is zero, since the model includes the intercept.

Indeed, $\underline{e}^T \underline{v} = 0$ for all $\underline{v} \in \mathcal{C}(X)$

$\sum_{i=1}^n e_i = \underline{e}^T \underline{1} = 0 \qquad \underline{1} \in \mathcal{C}(X).$

$\sum_{i=1}^n e_i x_{i2} = 0$ no

$\sum_{i=1}^n e_i \log(x_{i2}) = 0$ yes

$\sum_{i=1}^n e_i \log(x_{i2}^2) = \sum_{i=1}^n e_i [2 \log x_{i2}] = 2 \sum_{i=1}^n e_i \log x_{i2} = 0$ yes