Problem 1

Probab. representation of data:

(Y_i, Z_i, W_i), $i = 1, .., n$ random vectors distributed as a (generic) random vector (Y, Z, W).

Y = nitrogen concentration

Z = season $\in \{1, 2\}$
winter summer

W = soil type $\in \{1, 2, 3\}$
A B C

Linear model

- will parameterize $\mathbb{E}(Y | Z = g, W = h)$
$g = 1, 2$, $h = 1, 2, 3$

- will assume $var(Y | Z = g, W = h) = \sigma^2 (= const)$
$\forall g, h$

- no structure is assumed for $\mathbb{E}(Y | Z, W)$
  - $\rightarrow$ interaction model for two categorical covariates will be considered
  - $\rightarrow$ to obtain requested interpretation, sum contrasts will be used

$$\mathbb{C} := \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \qquad \mathbb{D} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} d_1^T \\ d_2^T \\ d_3^T \end{pmatrix}$$

Possible linear model; $g = 1, 2$, $h = 1, 2, 3$

$$\mathbb{E}(Y | Z = g, W = h) = \beta_0 + \beta^Z c_g + d_h^T \beta^W + c_g \cdot d_h^T \beta^{ZW},$$

$$\beta^W = (\beta_1^W, \beta_2^W)^T, \quad \beta^{ZW} = (\beta_1^{ZW}, \beta_2^{ZW})^T$$

with a chosen parameterization

$$\beta_0 = \overline{m}$$

$$\beta^z = \overline{m}_W - \overline{m} \qquad \left(\begin{array}{l} z = 1 \equiv \text{winter} \\ = 2 \equiv \text{summer} \end{array} \right)$$

$$-\beta^z = \overline{m}_S - \overline{m}$$

$$Y := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad Z := \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \qquad W := \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

model matrix (schematic)

$$X = \left(\begin{array}{cccccc} 1 & 1 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 1 & 1 & 0 & 1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 1 & 1 & -1 & -1 & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 1 & -1 & 1 & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 1 & -1 & 0 & 1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 1 & -1 & -1 & -1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right) \begin{array}{l} \text{season} \quad \text{soil type} \\ \text{winter, A} \\ \\ \text{winter, B} \\ \\ \text{winter, C} \\ \\ \text{summer, A} \\ \\ \text{summer, B} \\ \\ \text{summer, C} \\ \\ \end{array}$$

regression coefficients

$$\beta = (\beta_0, \beta^z, \beta^w_1, \beta^w_2, \beta^{zw}_1, \beta^{zw}_2)^T.$$

Problem 2

With chosen parameterization $\overline{m_W} = \beta_0 + \beta^Z$
$\overline{m_S} = \beta_0 - \beta^Z$

$\Rightarrow \theta = \overline{m_W} - \overline{m_S} = 2\beta^Z = l^T\beta$, where $l = (0,2,0,0,0,0)^T$

If each combination of season and soil type present in data, $\operatorname{rank}(X) = 6$ and $(X^TX)^{-1}$ exists.

$\hat{\beta} := (X^TX)^{-1}X^TY$ (LSE of β);

~~Xnw~~ $\hat{Y} := X\hat{\beta}$ (fitted values)

$U := Y - \hat{Y}$ (residuals)

$MS_e := \frac{1}{n-k}\|Y - \hat{Y}\|^2$, $k = 6$

$$T := \frac{l^T\hat{\beta} - l^T\beta}{\sqrt{MS_e\, l^T(X^TX)^{-1}l}}$$

new assumpt.

I will additionally assume $Y|X \sim N(X\beta, \sigma^2 I_n)$
$= Y|Z,W$

and will show that $T \sim t_{n-k}$.

(a) $\underline{l^T\hat{\beta} | X \sim N(l^T\beta, \sigma^2 l^T(X^TX)^{-1}l)}$ (for any $\beta \in \mathbb{R}^k$)
will be shown

$\mathcal{M}(X^T) = \mathbb{R}^k \ni l$, hence $\exists a: l = X^Ta$.

$\Rightarrow l^T\hat{\beta} = a^T \underbrace{X(X^TX)^{-1}X^T}_{=: H} Y = a^T\hat{Y}$

(hat matrix, projection matrix to $\mathcal{M}(X)$)

properties of H: $H^T = H$, $HH = H$

$l^T\hat{\beta} = \underbrace{a^T H}_{\text{function of } X} Y$ we assume $Y|X \sim N(X\beta, \sigma^2 I_n)$ 2

$$\Rightarrow l^T\hat{\beta}|X \sim N(a^T \underbrace{HX}_{X}\beta, \sigma^2 a^T \underbrace{HH^T}_{H} a)$$

$a^T X\beta = l^T\beta$

$a^T HH^T a = a^T X(X^TX)^{-1}X^T a = l^T(X^TX)^{-1}l$

That is, $l^T\hat{\beta}|X \sim N(l^T\beta, \sigma^2 l^T(X^TX)^{-1}l)$

$$\Rightarrow \text{given } X: \quad \frac{l^T\hat{\beta} - l^T\beta}{\sqrt{\sigma^2 l^T(X^TX)^{-1}l}} \sim N(0,1)$$

1b) $\frac{1}{\sigma^2}\|Y - \hat{Y}\|^2 \sim \chi^2_{n-k}$ (given X)

$\|Y - \hat{Y}\|^2 = \|\underbrace{(I - H)}_{=:M} Y\|^2 = \|MY\|^2$

$=: M$ (projection matrix to $M(X)^\perp$)

properties of M: $M^T = M$, $MM = M$,

$HM = 0$

M can also be written as $M = N \cdot N^T$,

where $N = (n_1, \dots, n_{n-k})$ is $n \times (n-k)$ matrix with an orthonormal vector basis of $M(X)^\perp$ in its columns, i.e.

~~NN^T~~ $N^TN = I_{n-k}$.

Hence $\|Y-\hat{Y}\|^2 = \|MY\|^2 = Y^T \underbrace{M^T M}_{M} Y = Y^T M Y =$

$= Y^T N N^T Y = \|\underbrace{N^T}_{\text{function of } X} Y\|^2.$

we assume $Y/X \sim N(X\beta, \sigma^2 I_n)$

$\Rightarrow \quad N^T Y / X \sim N(N^T X \beta, \sigma^2 N^T N)$

$N^T X = 0$ (scalar products of basis of $\mathcal{M}(X)^\perp$ with vectors from an orthogonal space $\mathcal{M}(X)$)

$N^T N = I_{n-k}$

That is $\quad N^T Y / X \sim N(0, \sigma^2 I_{n-k})$

$\Rightarrow \quad \frac{1}{\sigma} N^T Y / X \sim N(0, I_{n-k})$

properties of χ^2
$\Rightarrow \quad \|\frac{1}{\sigma} N^T Y\|^2 = \frac{1}{\sigma^2}\|N^T Y\|^2 \sim \chi^2_{n-k}$ (given X)

That is $\quad \frac{1}{\sigma^2}\|Y-\hat{Y}\|^2 \sim \chi^2_{n-k}$ (given X).

(c) $U = Y - \hat{Y}$ and $\hat{Y}$ are (given X) independent.

$\hat{Y} = H \cdot Y, \quad U = M \cdot Y$

Jointly: $\begin{pmatrix} \hat{Y} \\ U \end{pmatrix} = \begin{pmatrix} H \\ M \end{pmatrix} Y$ (linear fun. of N distributed rand. vector)

$\Rightarrow$ given X, $\begin{pmatrix} \hat{Y} \\ U \end{pmatrix} \sim N(\dots)$.

To show independence of $\hat{Y}$ and U, we only need to show that they are uncorrelated (independence then follows from properties of normal distribution):

$$\operatorname{cov}(\hat{Y}, U \mid X) = \operatorname{cov}(HY, MY \mid X) = H \cdot \operatorname{var}(Y \mid X) \cdot M^T =$$

$$= H(\sigma^2 I_n) M = \sigma^2 H \cdot M = 0$$

Now:

$$T = \frac{l^T\hat{\beta} - l^T\beta}{\sqrt{MSe\, l^T(X^TX)^{-1}l}} = \underbrace{\frac{l^T\hat{\beta} - l^T\beta}{\sqrt{\sigma^2 l^T(X^TX)^{-1}l}}}_{\sim N(0,1)} \sqrt{\frac{\sigma^2 (n-k)}{\|Y-\hat{Y}\|^2}}$$

given X: $\sim N(0,1)$ $\sim \chi^2_{n-k}$

moreover: $\|Y - \hat{Y}\|^2$ is a (quadratic) fun. of U

$l^T\hat{\beta}$ is a linear fun. of $\hat{Y}$,

hence (given X) $\|Y-\hat{Y}\|^2$ and $l^T\hat{\beta}$ are independent

definition of χ^2 distrib. $\Rightarrow$

$$T \mid X \sim t_{n-k}.$$

Distribution is the same for (almost) all values of the condition

$\Rightarrow$ also unconditionally $T \sim t_{n-k}$.

We have shown that $\forall \beta \in \mathbb{R}^k$ and $\forall \sigma^2 > 0$

$T \sim t_{n-k}$.

Hence $\forall \beta^0 \in \mathbb{R}^k$ $\forall \sigma_0^2 > 0$:

$$P\left(|T_0| < t_{n-k}\left(1-\tfrac{\alpha}{2}\right) ;\ \beta = \beta^0,\ \sigma^2 = \sigma_0^2 \right) = 1-\alpha,$$

where $$T_0 = \frac{\ell^T\hat{\beta} - \ell^T\beta^0}{\sqrt{MSe\ \ell^T(X^TX)^{-1}\ell}}$$

$$\Rightarrow P\left(\left(\ell^T\hat{\beta} \pm t_{n-k}\left(1-\tfrac{\alpha}{2}\right) \cdot \sqrt{MSe\ \ell^T(X^TX)^{-1}\ell}\right) \ni \ell^T\beta^0;\ \beta = \beta^0,\ \sigma^2 = \sigma_0^2 \right) = 1-\alpha$$

That is, interval $\left(\ell^T\hat{\beta} \pm t_{n-k}\left(1-\tfrac{\alpha}{2}\right)\sqrt{MSe\ \ell^T(X^TX)^{-1}\ell}\right)$ is the $(1-\alpha)100\%$ conf. interval for $\theta = \ell^T\beta$.

Problem 3

Let the following hold:

(A0) $(Y_1, X_1^T)^T, (Y_2, X_2^T)^T, \ldots \overset{iid}{\sim} (Y, X^T)^T$,

$\mathbb{E}(Y|X) = X^T\beta$, where $\beta \in \mathbb{R}^k$ is unknown

(A1) For $X = (X_0, \ldots, X_{k-1})^T$: $\mathbb{E}|X_j \cdot X_l| < \infty \ \forall j, l$

$\mathbb{E}XX^T = W > 0$

(A2) $\sigma^2(X) := \operatorname{var}(Y|X) = \sigma^2 = \text{const}$

$0 < \sigma^2 < \infty$ unknown.

Let $\hat{\beta}_n = (\mathbb{X}_n^T \mathbb{X}_n)^{-1}(\mathbb{X}_n^T \mathbb{Y}_n)$, where

$$\mathbb{X}_n = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}, \quad \mathbb{Y}_n = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ and}$$

$$MSe_n := \frac{1}{n-k} \sum_{i=1}^{n} (Y_i - X_i^T \hat{\beta}_n)^2.$$

Then $MSe_n \xrightarrow{a.s.} \sigma^2$ as $n \to \infty$.

Proof:

(i) Under given assumptions $\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T}_{\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n} \xrightarrow{a.s.} W$

~~and~~

and since $W > 0$, $\exists n_0 \ \forall n \geq n_0 \ \mathbb{X}_n^T \mathbb{X}_n > 0$

and also $n(\mathbb{X}_n^T \mathbb{X}_n)^{-1} \xrightarrow{a.s.} W^{-1} =: V$

at the same time $\forall n \geq n_0$, $\hat{\beta}_n$ is well defined and $\hat{\beta}_n \xrightarrow{a.s.} \beta$.

(ii) $MSe_{i,n} = \frac{1}{n-k} \sum_{i=1}^{n} (Y_i - X_i^T \hat{\beta}_n)^2 =$

$$= \underbrace{\frac{n}{n-k}}_{\downarrow n\to\infty \; 1} \cdot \frac{1}{n} \sum_i (Y_i - X_i^T \hat{\beta}_n)^2$$

Further, $\frac{1}{n} \sum_i (Y_i - X_i^T \hat{\beta}_n)^2 = \frac{1}{n} \sum_i (\underbrace{Y_i - X_i^T \beta}_{=: \varepsilon_i} + X_i^T \beta - X_i^T \hat{\beta}_n)^2 =$

$$= \underbrace{\frac{1}{n} \sum_i \varepsilon_i^2}_{A_n} + \underbrace{\frac{1}{n} \sum_i \{X_i^T (\beta - \hat{\beta}_n)\}^2}_{B_n} + \underbrace{\frac{2}{n} \sum_i (Y_i - X_i^T \beta) X_i^T (\beta - \hat{\beta}_n)}_{C_n}$$

(a) A_n:

$$\mathbb{E}(\varepsilon_i) = \mathbb{E}(\mathbb{E}(\varepsilon_i | X_i)) = \mathbb{E}(\underbrace{\mathbb{E}(Y_i - X_i^T \beta | X_i)}_{=0 \text{ by } (A0)}) = \mathbb{E}0 = 0$$

$$\mathbb{E}(\varepsilon_i^2) = \operatorname{var} \varepsilon_i = \mathbb{E}(\operatorname{var}(\varepsilon_i | X_i)) + \operatorname{var}(\mathbb{E}(\varepsilon_i | X_i)) =$$

$$= \mathbb{E}(\underbrace{\operatorname{var}(Y_i | X_i)}_{\sigma^2 \text{ by } (A2)}) + \operatorname{var}(0) = \mathbb{E}\sigma^2 = \sigma^2$$

by (A0) ε_i, $i = 1, 2, \ldots$ are iid and by SLLN: $A_n = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \xrightarrow{a.s.} \sigma^2$

(b) B_n:

$$B_n = \frac{1}{n} \sum_{i=1}^{n} (\beta - \hat{\beta}_n)^T X_i X_i^T (\beta - \hat{\beta}_n) =$$

$$= \underbrace{(\beta - \hat{\beta}_n)^T}_{\xrightarrow{a.s.} 0^T} \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T\right)}_{\xrightarrow{a.s.} W > 0} \underbrace{(\beta - \hat{\beta}_n)}_{\xrightarrow{a.s.} 0}$$

Hence $B_n \xrightarrow{a.s.} 0^T \cdot W \cdot 0 = 0$

(c) C_n:

$$C_n = \frac{2}{n} \sum_{i=1}^{n} (Y_i - X_i^T \beta) X_i^T (\beta - \hat{\beta}_n)$$

$$= 2 \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i X_i^T \right) \underbrace{(\beta - \hat{\beta}_n)}_{\downarrow a.s. \; 0}$$

The j-th element of vector $\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i X_i^T$ is

$$\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i X_{ij}$$

- $\varepsilon_i X_{ij}$ is a sequence of iid rand. variables by (A0)
- $\mathbb{E} |\varepsilon_i X_{ij}| \leq \sqrt{\underbrace{\mathbb{E} X_{ij}^2}_{<\infty \text{ by } A1} \cdot \underbrace{\mathbb{E} \varepsilon_i^2}_{<\infty \text{ by } A2}} < \infty$

$$\mathbb{E} \varepsilon_i X_{ij} = \mathbb{E}(\mathbb{E}(\varepsilon_i X_{ij} | X_i)) = \mathbb{E}(X_{ij} \underbrace{\mathbb{E}(\varepsilon_i | X_i)}_{=0}) = 0$$

Hence by SLLN: $\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i X_i^T \xrightarrow{a.s.} 0^T$

and also $C_n \xrightarrow{a.s.} 0$.

In summary $MSe_n = \frac{n}{n-k} (A_n + B_n + C_n)$

$$\begin{array}{cccc} \downarrow & \downarrow a.s. & \downarrow a.s. & \downarrow a.s. \\ 1 & \sigma^2 & 0 & 0 \end{array}$$

and hence $MSe_n \xrightarrow{a.s.} \sigma^2$.