4.4 Categorical covariate

$z_i \in \mathcal{Z} = \{1, \dots, G\}$, z_i categorical

just labels

$1 < \dots < G$ but ordering in general not important

Example: consumption ~ fdrive (plots) 77-78

regression function to be specified is

$$m(z) = E(Y/Z = z) \quad , \quad z = 1, \dots, G$$

$\rightarrow$ G conditional expectations

$$m(1) = E(Y/Z=1) =: m_1$$

$$\vdots \qquad\qquad \vdots$$

$$m(G) = E(Y/Z=G) =: m_G$$

---

consumption ~ fdrive 79

- empirical versions (estimates) of $m_1, \dots, m_G$

NOTATION: $m = (m_1, \dots, m_G)^T$

one-way classified group means

Appealing regression function?

$$m(z) = \beta_0 + \beta_1 z$$

$$\Longrightarrow \left.\begin{aligned} m_2 - m_1 &= \beta_1 \\ m_3 - m_2 &= \beta_1 \\ m_3 - m_1 &= 2\beta_1 \end{aligned}\right\} \text{ still holds if I change } \text{labels?}$$

4.4.1 Link to a G-sample problem

consumption ~ fdrive (boxplots)

Assume (without loss of generality):

Data $(Y_i; Z_i)^T$ sorted by Z values

$$\mathbb{Z} = \begin{pmatrix} z_1 \\ \vdots \\ z_{n_1} \\ \hline \vdots \\ \hline z_{n-n_G+1} \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \hline \vdots \\ \hline G \\ \vdots \\ G \end{pmatrix} \begin{array}{l} n_1\text{-times} \\ \\ \vdots \\ \\ n_G\text{-times} \end{array}$$

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ \hline \vdots \\ \hline Y_{n-n_G+1} \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \left.\begin{matrix} Y_{1,1} \\ \vdots \\ Y_{1,n_1} \end{matrix}\right\} \mathbb{Y}_1 \\ \hline \vdots \\ \hline \left.\begin{matrix} Y_{G,1} \\ \vdots \\ Y_{G,n_G} \end{matrix}\right\} \mathbb{Y}_G \end{pmatrix}$$

double subscript introduced

If additionally $(Y_i; Z_i)^T \overset{iid}{\sim} (Y, Z)$ with

$m_g = \mathbb{E}(Y/Z=g)$, $g = 1, \dots, G$, $\sigma^2 = var(Y/Z=g)$

Then

$$\left.\begin{matrix} Y_{1,1}, \dots, Y_{1,n_1} \overset{iid}{\sim} (m_1, \underline{\sigma}^2) \\ \vdots \\ Y_{G,1}, \dots, Y_{G,n_G} \overset{iid}{\sim} (m_G, \underline{\sigma}^2) \end{matrix}\right\} \text{homoscedastic}$$

- with random covariates $n_1, \dots, n_G$ are also random
- Assumption (for the rest of the lecture)

$$P(n_1 > 0) = \dots = P(n_G > 0) = 1$$

81

4.4.2 Linear model parameterization of one-way classified group means

$$\mathbb{E}(Y|Z) = \begin{pmatrix} m_1 \cdot 1\!\!1_{n_1} \\ \vdots \\ m_G \cdot 1\!\!1_{G} \end{pmatrix} =: \mu \underset{\text{to write it as}}{\overset{\text{We want}}{=}} X \beta$$

for some X and some β

regression space of a categorical covariate

$$= (\text{vector space}) \left\{ \begin{pmatrix} m_1 1\!\!1_{n_1} \\ \vdots \\ m_G 1_G \end{pmatrix} : m_1, \dots, m_G \in \mathbb{R} \right\} \subset \mathbb{R}^n$$

(space generated by vectors)

$$= \mathcal{M}_F(n_1, \dots, n_G)$$

$$n_1 > 0, \dots, n_G > 0 \implies \text{vec-dim}(\mathcal{M}_F) = G.$$

Possible (orthogonal) vector basis (possible $\underline{\underline{X}}$ matrix) is

$$Q = \begin{pmatrix} 1 & & 0 \\ \vdots & & \vdots \\ 1 & & 0 \\ \hline \vdots & & \vdots \\ \hline 0 & & 1 \\ \vdots & & \vdots \\ 0 & & 1 \end{pmatrix} = \begin{pmatrix} 1\!\!1_{n_1} \otimes (1, \dots, 0) \\ \vdots \\ 1\!\!1_{n_G} \otimes (0, \dots, 1) \end{pmatrix}$$

- with Q as a model matrix, $\mu = \mathbb{E}(Y|Z)$ would be parameterized as $\mu = Q \cdot \beta$, $\beta = (\beta_1, \dots, \beta_G)^T$ would be $\beta_j = m_j$, $j = 1, \dots, G$

<u>REMEMBER</u>: We want to write

$$\mathbb{E}(Y|Z) = \begin{pmatrix} m_1 \mathbb{1}_{n_1} \\ \vdots \\ m_G \mathbb{1}_{n_G} \end{pmatrix} \overset{\downarrow}{=} \mathbb{X}_{n \times k} \beta$$

- matrix $\mathbb{X}$ must satisfy: $k \geq G$, $\operatorname{rank}(\mathbb{X}) = G$
- $\mathbb{X}$ must be of the type

$$\mathbb{X} = \begin{pmatrix} x_1^T \\ \vdots \\ x_1^T \\ \hline \vdots \\ \hline x_G^T \\ \vdots \\ x_G^T \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{n_1} \otimes x_1^T \\ \vdots \\ \mathbb{1}_{n_G} \otimes x_G^T \end{pmatrix}$$

- $x_1, \dots, x_G \in \mathbb{R}^k$ chosen such that $\operatorname{rank}(\mathbb{X}) = G$

$$\tilde{X}_{G\times k} := \begin{pmatrix} x_1^T \\ \vdots \\ x_G^T \end{pmatrix}, \quad \text{clearly } \operatorname{rank}(\tilde{X}) = \operatorname{rank}(X) \underset{=}{\text{we want}} G$$

Once $\tilde{X}$ chosen and used within a linear model with ~~we~~ ~~have~~

$$\mu = E(Y|Z) = \begin{pmatrix} m_1 \mathbb{1}_{n_1} \\ \vdots \\ m_G \mathbb{1}_{n_G} \end{pmatrix} \downarrow = X\beta = \begin{pmatrix} x_1^T \\ \vdots \\ x_1^T \\ \hline \vdots \\ \hline x_G^T \\ \vdots \\ x_G^T \end{pmatrix} \beta,$$

we have ~~$m_1 =$~~ $E(Y|Z=1) = m_1 = x_1^T\beta$

$$\vdots$$

$$E(Y|Z=G) = m_G = x_G^T\beta$$

$$\equiv \quad \mathbb{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_G \end{pmatrix} = \tilde{X}\beta$$

One possible choice of $\tilde{X}$:

$\tilde{X} = I_G \quad (\operatorname{rank}(\tilde{X}) = G) \quad \Rightarrow \quad$ with $\beta = (\beta_1, \ldots, \beta_G)^T$.

$$= \begin{pmatrix} 1 & \cdots & 0 \\ 0 & \cdots & 1 \end{pmatrix} \qquad \begin{matrix} m_1 = \beta_1 \\ \vdots \\ m_G = \beta_G \end{matrix}, \quad \mathbb{m} = \beta$$

- it's perhaps surprising but $\tilde{X} = I_G$ leading to $m_g = \beta_g$, $g = 1, \dots, G$ is not always welcome
- the related model matrix

$$X = \left(\begin{array}{cc} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ \hline 0 & \\ \vdots & \\ 0 & \\ \hline 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{array}\right)$$

does not have an explicit intercept column

$\longrightarrow$ we will introduce such full-rank parameterizations that will lead to the model matrix

$$\left(\begin{array}{cc} 1 & \\ 1 & \\ \vdots & X \\ \vdots & \\ 1 & \end{array}\right)$$

4.4.3 Full-rank parameterization of one-way classified group means

We want to parameterize

$$\mathbb{E}(\mathbb{Y}|\mathbb{Z}) = \mu = \left(\begin{array}{c} m_1 \\ \vdots \\ m_1 \\ \hline \vdots \\ m_G \\ \vdots \\ m_G \end{array}\right) \stackrel{!}{=} \mathbb{X}\beta$$

$\mathbb{X}_{n\times G}$, $\text{rank}(\mathbb{X}) = G$

$$\beta = (\beta_0, \underbrace{\beta_1, \dots, \beta_{G-1}}_{\beta^Z})^T$$

We want an intercept column:

$$\mathbb{X} = \left(\mathbb{1} \;\; \begin{array}{c} c_1^T \\ \vdots \\ c_1^T \\ \hline \vdots \\ \hline c_G^T \\ \vdots \\ c_G^T \end{array}\right) = \left(\begin{array}{c} \mathbb{1}_{n_1} \otimes (1, c_1^T) \\ \vdots \\ \mathbb{1}_{n_G} \otimes (1, c_G^T) \end{array}\right)$$

It is sufficient to choose $\mathbb{C}_{G\times(G-1)} = \begin{pmatrix} c_1^T \\ \vdots \\ c_G^T \end{pmatrix}$

such that $\text{rank}((\mathbb{1}, \mathbb{C})) = G$

That is, we need:
- $\text{rank}(\mathbb{C}) = G-1$
- $\mathbb{1} \notin \mathcal{M}(\mathbb{C})$

(Columns of $\mathbb{C}$ linearly independent + linearly independent with $\mathbb{1}$)

with chosen $\mathbb{C}$:

$$\mathbf{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_G \end{pmatrix} \overset{\text{is parameterized as}}{=} \begin{pmatrix} 1 & \mathbf{c}_1^T \\ \vdots & \\ 1 & \mathbf{c}_G^T \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{G-1} \end{pmatrix} \Big\} \boldsymbol{\beta}^Z$$

that is as

$$m_1 = \beta_0 + \mathbf{c}_1^T \boldsymbol{\beta}^Z$$

$$\vdots$$

$$m_G = \beta_0 + \mathbf{c}_G^T \boldsymbol{\beta}^Z$$

$$\mathbf{m} = \beta_0 \cdot \mathbb{1} + \mathbb{C} \cdot \boldsymbol{\beta}^Z$$

Def 4.5 Full-rank parameterization of a categorical covariate — 83

Full-rank parameterization of a categorical covariate with G levels is a choice of the $G \times (G-1)$ matrix $\mathbb{C}$ that satisfies $\text{rank}(\mathbb{C}) = G-1$, $\mathbb{1}_G \notin \mathcal{M}(\mathbb{C})$.

---

often: columns of $\mathbb{C}$ are so called contrasts (vectors which elements sum up to zero).

$\mathbb{C}$: (pseudo)contrast matrix

Note (and REMEMBER):

Parameterization of a categorical covariate
≡ (general) definition of a covariate

We have defined (def 4.1) for a single covariate

$$\mathbb{E}(Y/Z=z) \overset{\text{parameterized as}}{=} \beta_0 + \underbrace{\mathbb{S}^T(z)}_{\text{parameterization of the covariate } Z}\beta^Z$$

NOW:
$$\mathbb{E}(Y/Z=z) \overset{\text{parameterized as}}{=} \beta_0 + \mathbb{C}_z^T\beta^Z$$

That is: $\boxed{\mathbb{C}_z = \mathbb{S}(z)}$

Corresponding reparameterizing matrix
(non-intercept columns in the model matrix):

$$\underset{\text{in general}}{\mathbb{S} = \begin{pmatrix} \mathbb{S}^T(Z_1) \\ \vdots \\ \mathbb{S}^T(Z_n) \end{pmatrix}} = \underset{\text{with categorical covariate}}{\begin{pmatrix} \mathbb{C}_{Z_1}^T \\ \vdots \\ \mathbb{C}_{Z_n}^T \end{pmatrix}}$$

appropriately recycled rows of the $\mathbb{C}$ matrix

Evaluation of the effect of the categorical covariate

$$E(Y/Z=z) = \beta_0 + \mathbb{C}_z^T \beta^z$$

$$\mathbb{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_G \end{pmatrix} = \beta_0 \cdot \mathbb{1}_G + \mathbb{C} \cdot \beta^z$$

$$E(Y/Z) = \begin{pmatrix} m_1 \\ \vdots \\ m_1 \\ \hline \vdots \\ \hline m_G \\ \vdots \\ m_G \end{pmatrix} = \mathbb{X} \cdot \beta = \underbrace{\begin{pmatrix} 1 & \mathbb{C}_1^T \\ \vdots & \vdots \\ 1 & \mathbb{C}_1^T \\ \hline \vdots & \vdots \\ \hline 1 & \mathbb{C}_G^T \\ \vdots & \vdots \\ 1 & \mathbb{C}_G^T \end{pmatrix}}_{\substack{G \text{ columns} \\ \text{rank} = G}} \begin{pmatrix} \beta_0^z \\ \beta^z \end{pmatrix}$$

$H_0: m_1 = \cdots = m_G \quad \equiv H_0: \beta^z = \mathbb{0}$

$\equiv H_0: E(Y|Z) \in \mathcal{M}(\mathbb{1})$ (submodel)

→ Wald type test (F-test) on a subvector of regression coefficients

≡ submodel F-test

- $G = 2$: → (equal variances) two-sample t-test
- $G > 2$: → one-way ANOVA F-test

Example: Cars2004nh

Y = consumption, Z = drive $\in \{1, 2, 3\}$

1 = front, 2 = rear, 3 = 4x4

Easy homework:

$$\hat{Y} = \begin{pmatrix} \bar{Y}_1 \\ \vdots \\ \hline \bar{Y}_2 \\ \vdots \\ \hline \bar{Y}_3 \\ \vdots \end{pmatrix}$$

Since $\hat{Y} = X \cdot \hat{\beta}$ and X has a full-rank

then $\hat{\beta}$ = linear function of $\hat{Y}$

(see also Gauss-Markov th.)

Some choices of $\mathbb{C}$ matrices will follow

+ discussion of implied interpretation

of β's.

Reference group pseudocontrasts (dummies)

86

contr. treatment in R

$$\mathbb{C} = \begin{array}{c} \begin{pmatrix} 0 & \cdots & 0 \\ 1 & \cdots & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \end{array} \begin{array}{l} 1 \\ 2 \\ \vdots \\ G \end{array} = \begin{pmatrix} 0_{G-1}^T \\ \\ I_{G-1} \end{pmatrix} \begin{array}{l} \leftarrow \text{row in model matrix for units with } Z=1 \\ \leftarrow \text{with } Z=2 \\ \vdots \\ \leftarrow \text{with } Z=G \end{array}$$

$$\mathbb{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_G \end{pmatrix} = \beta_0 \cdot \mathbb{1} + \mathbb{C} \cdot \beta^Z \quad , \quad \beta^Z = (\beta_1, \ldots, \beta_{G-1})^T$$

$$\Rightarrow \quad \begin{aligned} m_1 &= \beta_0 \\ m_2 &= \beta_0 + \beta_1 \\ &\vdots \\ m_G &= \beta_0 + \beta_{G-1} \end{aligned} \qquad \Rightarrow \quad \begin{aligned} \beta_0 &= m_1 \\ \beta_1 &= m_2 - m_1 \\ &\vdots \\ \beta_{G-1} &= m_G - m_1 \end{aligned}$$

$$E(Y|Z=z) = m(z) = \beta_0 \mathbb{1} + \beta_1 \underbrace{\mathbb{I}(z=2)}_{X_1} + \ldots + \beta_{G-1} \underbrace{\mathbb{I}(z=G)}_{X_{G-1}}$$

$$X^T = (X_0 \,,\, X_1 \,,\, \ldots \,,\, X_{G-1})$$

vector of regressors

$\equiv$ dummy variables

R output (Cars 2004nh)

87

- any group can be chosen as reference
- e.g. <u>contr. SAS</u> in R

$$\mathbb{C} = \begin{pmatrix} 1 & \dots & 0 \\ & \ddots & \\ 0 & & 1 \\ 0 & \dots & 0 \end{pmatrix}, \quad m = \beta_0 \cdot \mathbb{1} + \mathbb{C} \cdot \beta^z$$

$$\beta^z = (\beta_1, \dots, \beta_{G-1})^T$$

$$\Rightarrow \quad m_1 = \beta_0 + \beta_1 \qquad\qquad \beta_0 = m_G$$

$$\vdots \qquad \Rightarrow \qquad \beta_1 = m_1 - m_G$$

$$m_{G-1} = \beta_0 + \beta_{G-1} \qquad\qquad \vdots$$

$$m_G = \beta_0 \qquad\qquad \beta_{G-1} = m_{G-1} - m_G$$

<u>R output / (Cars 2004nh)</u>

Sum contrasts

contr.sum in R

$$\mathbb{C} = \begin{array}{c} \begin{pmatrix} 1 & \cdots & 0 \\ & & \\ 0 & \cdots & 1 \\ -1 & \cdots & -1 \end{pmatrix} \end{array} \begin{array}{c} 1 \\ \\ G-1 \\ G \end{array} = \begin{pmatrix} I_{G-1} \\ -\mathbb{1}_{G-1}^T \end{pmatrix}$$

row in model matrix for units with $z=1$

with $z=G-1$

with $z=G$

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_G \end{pmatrix} = \beta_0 \mathbb{1} + \mathbb{C}\beta^z \quad , \quad \beta^z = (\beta_1, \dots, \beta_{G-1})^T$$

$$\Rightarrow \boxed{\begin{aligned} m_1 &= \beta_0 + \beta_1 \\ &\vdots \\ m_{G-1} &= \beta_0 + \beta_{G-1} \\ m_G &= \beta_0 - \sum_{g=1}^{G-1} \beta_g \end{aligned}}$$

$$\sum_{g=1}^{G} m_g = G \cdot \beta_0$$

$$\beta_0 = \frac{1}{G} \sum_{g=1}^{G} m_g =: \bar{m}$$

$$\beta_1 = m_1 - \bar{m}$$

$$\beta_{G-1} = m_{G-1} - \bar{m}$$

$$E(Y \mid Z = z) = m(z) = \beta_0 + \beta_1 \mathbb{I}(z=1) + \dots + \beta_{G-1} \mathbb{I}(z=G-1) - \left(\sum_{g=1}^{G-1} \beta_g\right) \cdot \mathbb{I}(z=G)$$

$\rightarrow$ not really interesting

can be viewed as (less-than-full-rank) parameterization

$$m_g = \alpha_0 + \alpha_g \quad , g = 1, \dots, G$$

$\alpha = (\alpha_0, \alpha_1, \dots, \alpha_G)^T$: vector of parameters if we define $\alpha_1 = \beta_1 \quad (= m_1 - \bar{m})$

$\vdots$

$\alpha_{G-1} = \beta_{G-1} \quad (= m_{G-1} - \bar{m})$

$\alpha_G = -\sum_{g=1}^{G-1} \beta_g \quad (= m_G - \bar{m})$

the α's then satisfy $\boxed{\sum_{g=1}^{G} \alpha_g = 0}$

so called identifying constraint

HERE: "sum is zero"

→ sum contrasts

REMARK: least squares estimation

$$\hat{\alpha}_1 = \hat{\beta}_1, \ldots, \hat{\alpha}_{G-1} = \hat{\beta}_{G-1}$$

$$var(\hat{\alpha}_1 | Z) = \sigma^2 (X^T X)^{-1}_{\cdot\cdot} \qquad var(\hat{\alpha}_{G-1} | Z) = \sigma^2 (X^T X)^{-1}_{\cdot\cdot}$$

$$\hat{\alpha}_G = -\sum_{g=1}^{G-1} \hat{\beta}_g$$

$$var(\hat{\alpha}_G | Z) = \sigma^2 \cdot l^T (X^T X)^{-1} l, \quad l = (0, -1, \ldots, -1)^T$$

R output (Cars2004nh) 91

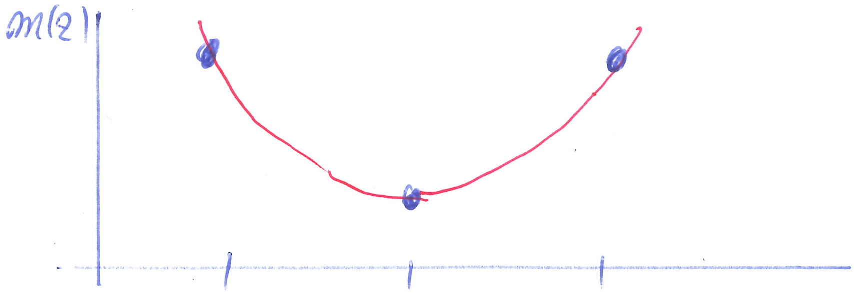
Orthonormal polynomial contrasts [92]

Example: Cars2004nh — contr.poly in R

Y = consumption, Z = categorized weight $\in 1 < 2 < 3 < 4 < 5$

- ordinal covariate

PLOT: (sample) means [93]



[94]

$\omega_1 < \omega_2 < \omega_3$ ← chosen labels

1 2 3

G numbers $(m_1, \dots, m_G)$ can ALWAYS be interpolated by a polynomial of degree $G-1$

→ why not to parameterize $m(z)$ as a polynomial of degree $G-1$

$$m_1 = m(\omega_1) = \beta_0 + \beta_1 P^1(\omega_1) + \dots + \beta_{G-1} P^{G-1}(\omega_1)$$ [95]

$$m_G = m(\omega_G) = \beta_0 + \beta_1 P^1(\omega_G) + \dots + \beta_{G-1} P^{G-1}(\omega_G)$$

$P^j(z)$: (orthonormal) polynomial of degree j

Let $w_1 < \dots < w_G$ be
an equidistant sequence of the group labels

$$P^j(z) = a_{j0} + a_{j1} z + \dots + a_{jj} z^j \quad , \quad j = 1, \dots, G-1$$

$\equiv$ orthonormal polynomial of degree j built above a sequence of the group labels (as in the case of numeric covariate)

$$\rightarrow \quad \mathbb{C} = \begin{pmatrix} P^1(w_1) & \dots & P^{G-1}(w_1) \\ \vdots & & \vdots \\ P^1(w_G) & & P^{G-1}(w_G) \end{pmatrix}$$

NOTE: Up to an orientation, columns of $\mathbb{C}$ are (for given G) invariant (up to orientation) towards the choice of equidistant group labels $w_1, \dots, w_G$.

$\mathbb{C}$ matrices for $G = 2, 3, 4$

REMEMBER interpretation:

$m_1 = E(Y|Z=1) = \beta_0 + \beta_1 P^1(w_1) + \cdots + \beta_{G-1} P^{G-1}(w_1)$

$\vdots$

$m_G = E(Y|Z=G) = \beta_0 + \beta_1 P^1(w_G) + \cdots + \beta_{G-1} P^{G-1}(w_G)$

Example: Cars2004nh

R output (F-statistic?) 99

contr.treatment

R output (F-statistic?) 100

contr.poly

101

Is it possible to consider the ordinal covariate as numeric having values 1, 2, 3, 4, 5 while considering only the cubic (degree = 3) polynomial trend $E(Y|Z=g) = \beta_0 + \beta_1 g + \beta_2 g^2 + \beta_3 g^3$?

$(+\cancel{\beta_4 g^4})$

- remind the same P-value for terms mweight^4 now and fweight^4 in previous slide

PLOT: by eyes

Is a linear trend that assumes

$$\mathbb{E}(Y|Z=g) = \beta_0 + \beta_1 g$$

adequate for data at hand?

M_0: $\mathbb{E}(Y|Z=g) = \beta_0 + \beta_1 g$

M_1: $\mathbb{E}(Y|Z=g) = m_g$ → parameterized arbitrarily

⇒

M_0 is submodel of M_1

→ F-test

HERE: linear trend rejected