5.5 Multiple regression model

- strategies to combine various covariates in a model
- ~~not yet~~ model building strategies

5.5.1 Model terms

Most applications of linear model:
each covariate $z \in \mathcal{Z} \subseteq \mathbb{R}$ enters the model

(a) using certain parameterization

$$s(z) = (s_1(z), \dots, s_{k-1}(z))^T \leftarrow \text{parameterization}$$

$$S = \begin{pmatrix} s^T(z_1) \\ \vdots \\ s^T(z_n) \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \leftarrow \text{reparameterizing matrix}$$

regressors derived from covariate values $z_1, \dots, z_n$

(b) inside interaction

(c) inside higher-order interaction

SUMMARY OF PARAMETERIZATIONS (covered by us)

79

Z numeric

(i) simple transformation

$$\mathbb{S}(z) = \underbrace{s(z)}_{t(z)} \qquad \mathbb{S} = \underbrace{\begin{pmatrix} s(z_1) \\ \vdots \\ s(z_n) \end{pmatrix}}_{T}$$

80

(ii) polynomial parameterization

$$\mathbb{S}(z) = (s_1(z), \dots, s_{k-1}(z))^T$$

$$= (\underbrace{p^1(z)}_{t(z)}, \underbrace{\dots}_{t^2(z)}, \underbrace{p^{k-1}(z)}_{t^{k-1}(z)})^T$$

$$\mathbb{S} = \begin{pmatrix} p^1(z_1) & \dots & p^{k-1}(z_1) \\ \vdots & & \vdots \\ p^1(z_n) & & p^{k-1}(z_n) \end{pmatrix}$$

$$\underbrace{}_{T} \quad T^2 \quad T^{k-1}$$

81

(iii) spline parameterization

$$\mathbb{S}(z) = (s_1(z), \dots, s_k(z))^T$$

$$= \underbrace{(B_1(z), \dots, B_k(z))^T}_{t(z)}$$

$$\mathbb{S} = \underbrace{\begin{pmatrix} B_1(z_1) & \dots & B_k(z_1) \\ \vdots & & \vdots \\ B_1(z_n) & & B_k(z_n) \end{pmatrix}}_{T}$$

Z categorical, $z \in \{1, .., G\}$

82

$$\mathbb{S}(z) = \underbrace{\mathbb{C}_z}_{t(z)} \qquad \mathbb{S} = \underbrace{\begin{pmatrix} \mathbb{C}_{z_1}^T \\ \vdots \\ \mathbb{C}_{z_n}^T \end{pmatrix}}_{T} \; (G-1) \text{ columns}$$

chosen (pseudo) contrast matrix

$$\mathbb{C} = \begin{pmatrix} \mathbb{C}_1^T \\ \vdots \\ \mathbb{C}_G^T \end{pmatrix}$$

(G−1) cols.

Def 5.3 The main effect model term

83

Depending on a chosen parameterization $\mathfrak{s}: \mathcal{Z} \to \mathbb{R}^{k^*}$, the main effect model term (of order one) of a given covariate Z is defined as

a transformation $\underline{t}$ with elements as follows

and a matrix $\overline{\overline{A}}$ with columns as follows.

→ page 54

83 - 86

Def 5.4 The main effect model term of order j

87

If a numeric covariate Z ~~is~~ parameterized using the polynomial of degree $k-1$, i.e., $\mathfrak{s} = (s_1, \dots, s_{k-1})^T$, $s_j(z) = P^j(z)$, $j = 1, \dots, k-1$, then the main effect model term of order j means

......

→ page 54

Covariates Z $\rightarrow$ main effect model terms
t_Z, $\mathbb{T}_Z$, $t_Z = (t_Z^1, \dots, t_Z^{k-1})^T$
W $\rightarrow$ t_W, $\mathbb{T}_W$, $t_W = (t_W^1, \dots, t_Z^{\ell-1})^T$ 88

Def. 5.5 The two-way interaction model term

The two-way interaction model term means elements of a vector $t_W \otimes t_Z$ and a matrix $\mathbb{T}^{ZW}$, where $\mathbb{T}^{ZW} := \mathbb{T}_Z : \mathbb{T}_W$.

Reminder:

$$(t_W \otimes t_Z)^T = (t_Z^1 \cdot t_W^1, \dots t_Z^{k-1} \cdot t_W^1, \dots, t_Z^1 \cdot t_W^{\ell-1}, \dots, t_Z^{k-1} \cdot t_W^{\ell-1})^T$$

- The main effect model term $\mathbb{T}_Z$ and/or the main effect model term $\mathbb{T}_W$ that enters the two-way interaction may also be of a degree $j > 1$.
- Both the main effect model terms $\mathbb{T}_Z$ and $\mathbb{T}_W$ are called as lower order terms included in the two-way interaction term $\mathbb{T}_Z : \mathbb{T}_W$.

Covariates and their main effect model terms 89

$Z \rightarrow t_Z, \mathbb{T}_Z$

$W \rightarrow t_W, \mathbb{T}_W$

$V \rightarrow t_V, \mathbb{T}_W$

Def 5.6 The three-way interaction model term

The three-way interaction model term means a vector $t_V \otimes (t_W \otimes t_Z)$ and a matrix $\mathbb{T}^{ZWV}$, where $\mathbb{T}^{ZWV} := (\mathbb{T}_Z : \mathbb{T}_W) : \mathbb{T}_V$.

+ Notes

5.5.2 Model Formula

≡ concise description of the model / model matrix

→ slide

e.g. $Y \sim 1 \quad \equiv \quad \mathbb{E}(Y|Z) = \beta_0$

$$X = \mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$Y \sim z + \log(w) + z : \log(w)$

$$\equiv \mathbb{E}(Y|Z=z, W=w) =$$

$$= \beta_0 + \beta_1 z + \beta_2 \log(w) + \beta_3 z \cdot \log(w)$$

$$X = \begin{pmatrix} 1 & z_1 & \log(W_1) & z_1 \cdot \log(W_1) \\ \vdots & \vdots & \vdots & \vdots \\ \underline{1} & \underline{z_n} & \underline{\log(W_n)} & \underline{z_n \cdot \log(W_n)} \end{pmatrix}$$

intercept | main effect of z | main effect of W | interaction

5.5.3 Hierarchically well formulated model

91

Def 5.7 Hierarchically well formulated model

Hierarchically well formulated (HWF) model is such a model that contains an intercept term (possibly implicitely) and with each model term also all lower order terms that are nested in this term.

Reason: Invariance of the regression space towards linear transformations of regressors.

92

Example: covariates $X_1, \dots, X_n$, $n > 3$,
at least three different values among $X_1, \dots, X_n$

quadratic regression function

HWF: $m_X(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, $\boxed{\text{rank} = 3}$

transformation $x \to t$ $(\delta \neq 0, \varphi \neq 0)$

$$x = \delta(t - \varphi), \quad t = \varphi + \frac{x}{\delta}$$

in t parameterization

$$m_t(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2, \quad \boxed{\text{rank} = 3}$$

$$\gamma_0 = \beta_0 - \beta_1 \delta \varphi + \beta_2 \delta^2 \varphi^2$$

$$\gamma_1 = \beta_1 \delta - 2\beta_2 \delta^2 \varphi$$

$$\gamma_2 = \beta_2 \delta^2$$

quadratic regression function, no linear term

- non HWF

$$m_x(x) = \beta_0 + \beta_2 x^2 \qquad \boxed{\text{rank} = 2}$$

transformation (as before):

$$x \to t \qquad (\delta \neq 0,\ \varphi \neq 0)$$

$$x = \delta(t - \varphi), \quad t = \varphi + \frac{x}{\delta}$$

in t parameterization

$$m_t(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 \qquad \boxed{\text{rank} = 3}$$

$$\gamma_0 = \beta_0 + \beta_2 \delta^2 \varphi^2$$

$$\gamma_1 = -2 \beta_2 \delta^2 \varphi$$

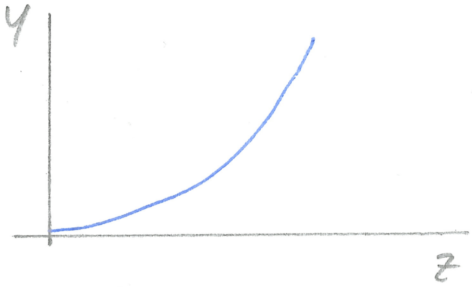
$$\gamma_2 = \beta_2 \delta^2$$

Possible reasons for not using the HWF model

- **no intercept**

e.g. $E(Y|Z=z) = \beta_1 z + \beta_2 z^2$

$\equiv$ ASSUMPTION $E(Y|Z=0) = 0$

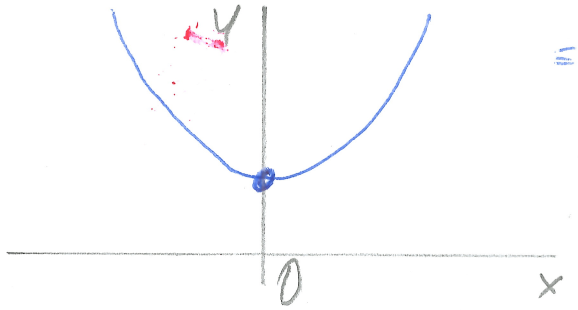


e.g. Z = velocity

Y = breaking distance

- **no linear term**

e.g. $E(Y|X=x) = \beta_0 + \beta_1 x^2$



$\equiv$ ASSUMPTION that the main vertex of a parabole goes throug 0 in X parameteriz.

- **no main effect**

e.g. $E(Y|X=x, Z=z) = \beta_0 + \beta_1 z + \beta_2 x \cdot z$

$\equiv E(Y|X=x, Z=0) = \beta_0$.

$\equiv$ ASSUMPTION that if $Z=0$, then $E(Y|\ldots)$ does not depend on X.

5.5.4 Usual strategy to specify a multiple regression model

Mostly, the linear model is based on

- main effect terms
- two-way interactions

- higher order interactions are only rarely considered

INTERPRETATION:

- covariate Z not included in any interaction with remaining covariates W

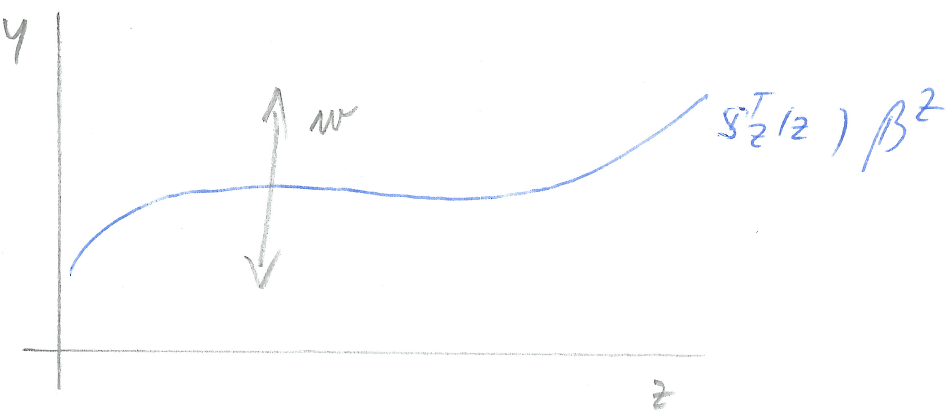
if Z parameterized by S_Z, the model is

$$E(Y \mid Z=z, W=w) = \beta_0 + S_Z^T(z)\beta^Z + m_W(w)$$

$$= \underbrace{(\beta_0 + m_W(w))}_{} + S_Z^T(z)\beta^Z$$

W modifies only intercept of Y–Z association

- Z has an additive effect in a model with all covariates



given $W=w$, interpretation of the model is as in sections 5.2 - 5.4 under additivity

Example: Cars2004nh, $n=409$

Y = consumption

covariates: drive (categorical) $z \in \{1,2,3,4\}$

engine size $\longrightarrow v$

weight $\longrightarrow \log(w)$

drop1: F-tests on submodels obtained by removing one term from the model while keeping the submodel still HWF

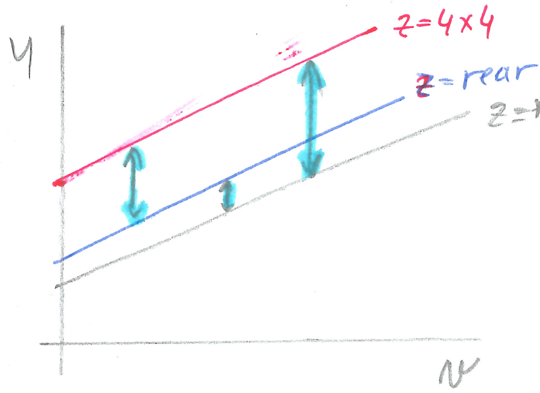
parameterized through reference group (pseudo)contrasts

$$\mathbb{C} = \begin{array}{cc} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} & \begin{array}{l} \text{front} \\ \text{rear} \\ 4\times 4 \end{array} \end{array}$$

$(\mathbb{I}(\text{rear}), \mathbb{I}(4\times 4))$ = regressor

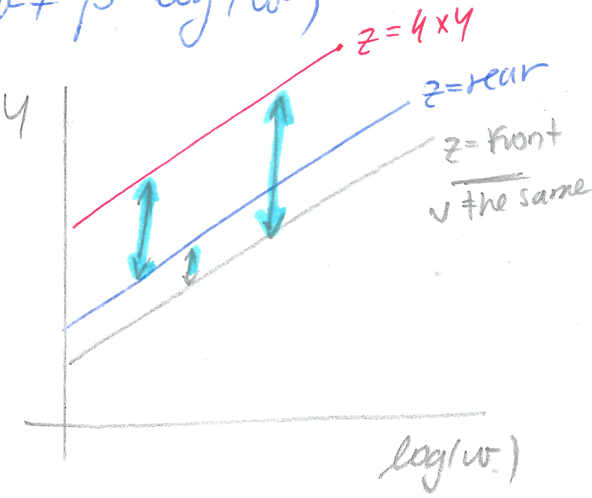
ADDITIVE MODEL

$$E(Y | Z=z, V=v, W=w) = \beta_0 + \mathbb{C}_z^T \beta^Z + \beta^v v + \beta^w \log(w)$$



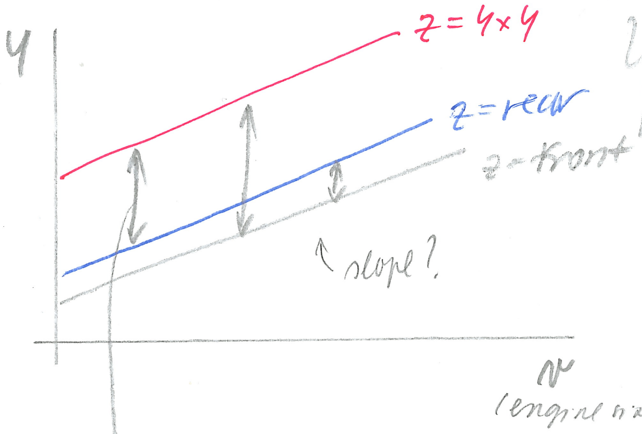
w the same

the same as well (on left and right)



+ drive: log(weight) interaction

- engine size still additive

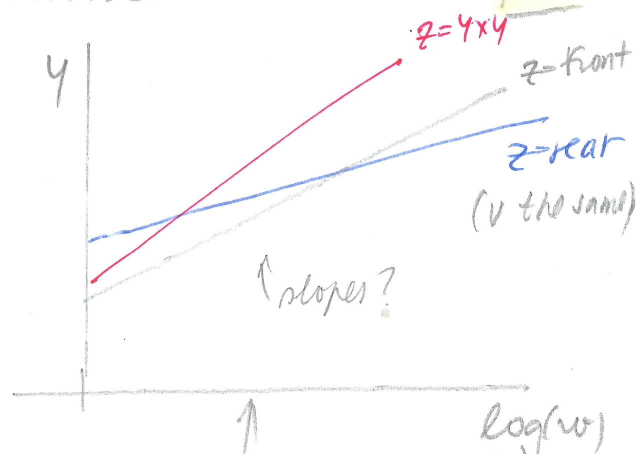


w the same

change of w only

each line slope remains the same

those differences now depend on w (weight) chosen to prepare plot



For given w and v differences between "z" groups depend on w

INTERPRETATION, cont'd

- covariates Z, W involved in an interaction

$\Rightarrow \boxed{Z \text{ and } W \text{ are } \cancel{\text{mut}} \text{ mutual effect modifiers}}$

Z parameterized by S_Z

W parameterized by S_W

$$\mathbb{E}(Y \mid Z=z, W=w, \underbrace{V=v}_{\text{remaining covariates}}) = \beta_0 + S_Z^T \beta^Z + S_W^T \beta^W + (S_W^T \otimes S_Z^T)\beta^{ZW} + m_{Z,V}(z,v) + m_{W,V}(w,v)$$

given $V=v$: $m_{Z,V}(z,v) = \gamma_0(v) + S_Z^T \gamma^Z(v)$

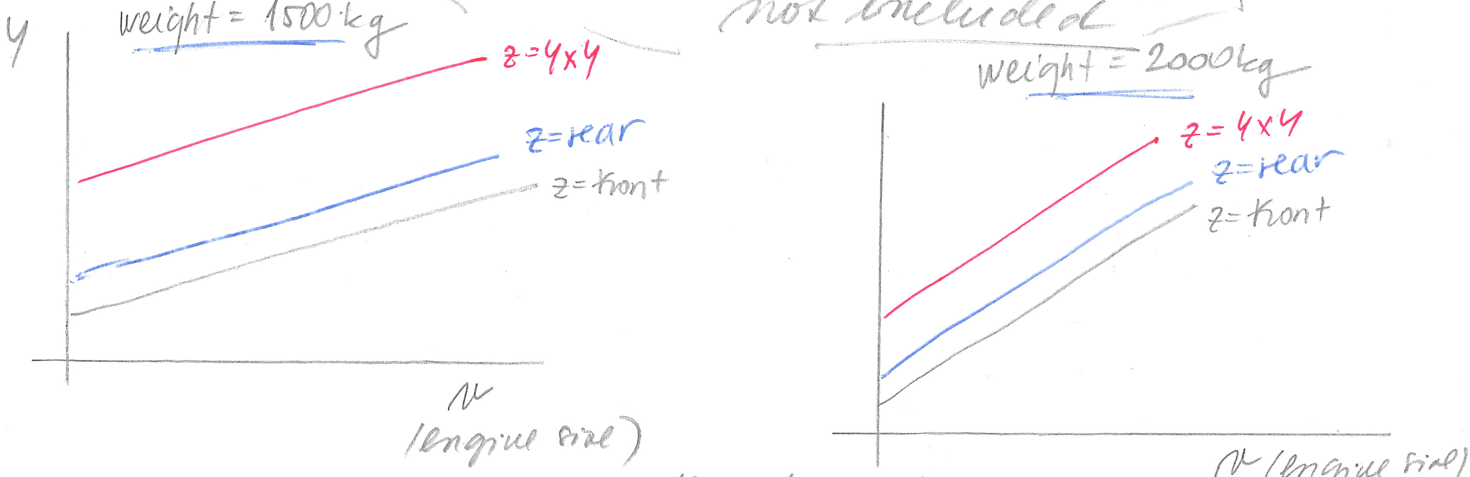
$m_{W,V}(w,v) = \delta_0(v) + \underbrace{S_W^T \delta^W(v)}_{\text{this part included if also } Z:V \text{ or } W:V \text{ interact in the model}}$

$$\Rightarrow E(Y \mid Z=z, W=w, V=v) = (\beta_0 + \gamma_0(v) + \delta_0(v)) + S_Z^T(\beta^Z + \gamma^Z(v)) + S_W^T(\beta^W + \underbrace{\delta^W(v)}_{\text{only if } Z:V \text{ or } W:V \text{ interact in the model}}) + (S_W^T \otimes S_Z^T)\beta^{ZW}$$

$\Rightarrow$ given $V=v$: interpretation of effect of Z and W on $E(Y \mid \cdots)$ is the same as in Sec. 5.2–5.4 (interaction pieces)

consumption ~ drive + engine size + log(weight) 99
+ drive: log(weight) + engine size: log(w)

drive: engine size interaction not included



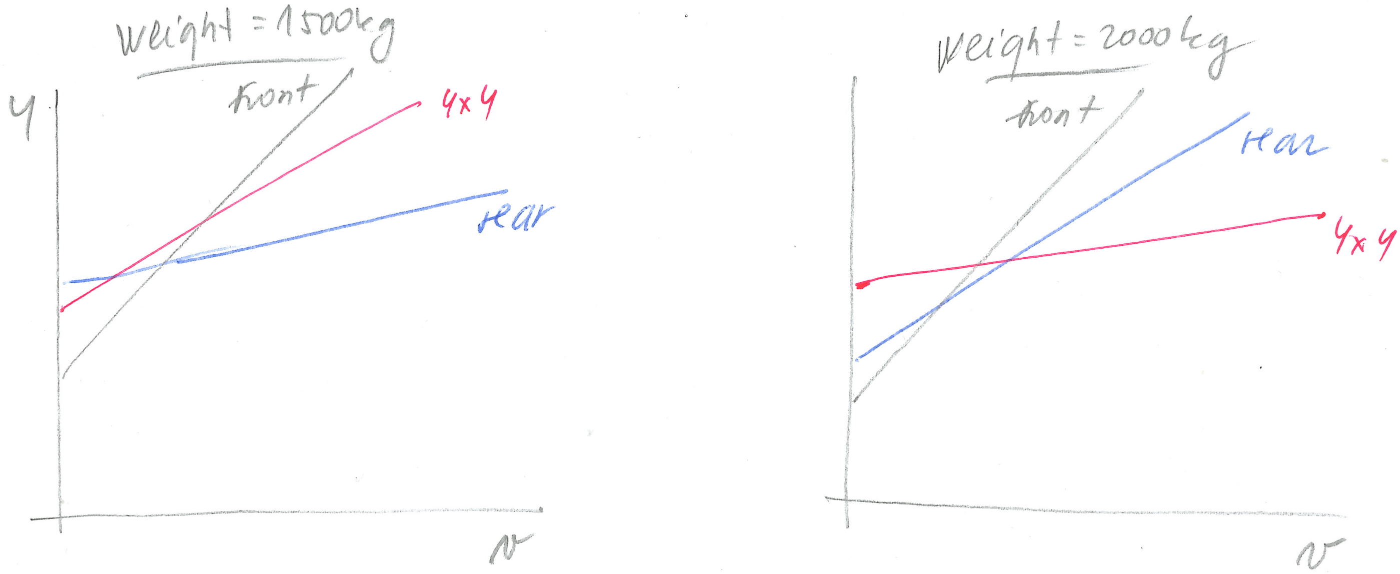
different weights

change of slope

---

consumption ~ (drive + engine size + log(weight))^2 100

-all interactions included



---

enlightenment (osvěta) 101 - 105

- residuals vs. fitted values may not show a problem
- residuals vs. regressors should also be used!

Comparison of considered models by F-tests 106

5.5.5 ANOVA tables

Certain model M ⟶ ANOVA table

= ~~ce~~ input to certain sequence of F-tests on submodels

Effect (term)	Degrees of freedom	Effect sum of squares	Effect mean square	F	P-value
*	*	*	*	*	*
*	*	*	*	*	*
Residual	ν_e	SS_e	MS_e		

↗ From model M for which ANOVA table is produced

denominator for all F-tests

⟶ Illustrations for the model

$M \equiv M_{AB}: \quad \sim A + B + A:B$ will follow.

A ↗ drive; B ↗ log(weight)

Cars2004nh (n = 409)

<u>IN GENERAL:</u>

<u>Rows of the table</u>: input to test $\underbrace{M_1 \subset M_2}$ $\subseteq M$

differ on rows and for tables of different type

<u>Effect</u>:

Degrees of freedom $\quad \nu_E = \nu_1 - \nu_2$

Sum of squares $\quad SS_E = SS(M_2 \mid M_1) = SS_e^1 - SS_e^2$

Mean square $\quad MS_E = \dfrac{SS_E}{\nu_E}$

F-statistic $\quad \dfrac{MS_E}{MS_e} \overset{H_0: M_1}{\sim} F_{\nu_E, \nu_e}$

Consider a model $M_{AB} : \sim A + B + A{:}B$ 108

- possible HWF submodels

$M_0 : \sim 1$

$M_A : \sim A$

$M_B : \sim B$

$M_{A+B} : \sim A + B$

$$M_0 \begin{array}{c} \subset M_A \subset \\ \subset M_B \subset \end{array} M_{A+B} \subset M_{AB}$$

$\boxed{\text{Type I (sequential) ANOVA table}}$ 109

Row E : M_1 (submodel) $\equiv$ all terms from previous rows

M_2 (bigger model) $\equiv M_1 + E$

- note the difference in the effect sum of squares on rows A and B in the two tables 110

---

summary to read 111

It also holds that $\sum_{\text{rows of table}} SS_E = SS_T$

$$\text{basis of } \mathbb{R}^n: \ (\mathbb{1} \ \overbrace{\underbrace{Q_A}_{SS(A|1)} \ \underbrace{Q_B}_{SS(A+B|A)} \ \underbrace{Q_{AB}}_{SS(AB|A+B)} \ \underbrace{N}_{SSe}}^{SS_T = SSe^0})$$

P-values: "sequential" building of the model

$1 \to A \to A+B \to A+B+A{:}B$

$1 \to B \to A+B \to A+B+A{:}B$

R software: anova(m)

---

illustration 112-114

Type II table

115

Row E : M_1 (submodel) = Full model − E
− all higher order terms than E that include E

M_2 (bigger model) = M_1 + E

- all compared models are HWF

---

summary to read 116

R software : Anova(m, type="II")
from package 'car'

---

illustration 117-118

Type III table

119

Row E : M_1(submodel) = Full model − E

terms in a model corresponding to E in given parameterization.

M_2 = Full model M

- not all compared models $M_1 \subset M_2$ are HWF

---

summary to read

120

R software: Anova(m, type = "III")

from package 'car'

---

contr. treatment used for drive

121

reference = Front

row lweight F-test ≡ slope for front = 0 ?

Contr. SAS used for drive

122

reference = 4x4

row lweight F-test ≡ slope for 4x4 = 0 ?

---

contr. sum used for drive

123

remember: $\beta^{lweight}$ = mean slope

row lweight F-test ≡ mean of the three slopes = 0 ?